Analyzing the neural transcriptional landscape in time and space

Dissertation

zur Erlangung des akademischen Grades

Doctor rerum naturalium

(Dr. rer. nat.)

im Fach Biologie

eingereicht an der

Lebenswissenschaftlichen Fakultät

der Humboldt-Universität zu Berlin

von

M.Sc. Christin Sünkel

geboren am xx.xx.xxxx in xxxxxx als Christin Stottmeister

Präsidentin der Humboldt-Universität zu Berlin

Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Lebenswissenschaftlichen Fakultät

Prof. Dr. Bernhard Grimm

Gutachter/innen:

1. Prof. Dr. Nikolaus Rajewsky
2. Prof. Dr. Markus Landthaler
3. Sebastian Kadener, PhD

Tag der mündlichen Prüfung: 02.12.2019

“Der Mensch muss das Gute und Große wollen.

Das Übrige hängt vom Schicksal ab.”

Alexander von Humboldt

“Bilde dich selbst, und dann wirke auf andere
durch das, was du bist.”

Wilhelm von Humboldt

Abstract

Circular RNAs (circRNAs) are an endogenous class of animal RNAs. Despite their abundance, their function and expression in the nervous system are unknown. Therefore, a circRNA catalogue comprising RNA-seq samples from different brain regions, primary neurons, synaptoneurosomes, as well as during neuronal differentiation was created. Using these and other available data, thousands of neuronal human and mouse circRNAs were discovered and analyzed. CircRNAs were extraordinarily enriched in the mammalian brain, well conserved in sequence, often expressed as circRNAs in both human and mouse, and sometimes even detected in *Drosophila* brains. CircRNAs were overall upregulated during neuronal differentiation, highly enriched in synapses, and often differentially expressed compared to their corresponding mRNA isoforms. CircRNA expression correlated negatively with expression of the RNA-editing enzyme ADAR1. Knockdown of ADAR1 induced elevated circRNA expression. Together, a circRNA brain expression atlas and evidence for important circRNA functions is provided.

Starting from this catalogue a circRNA, circSLC45A4 was identified. It is the main RNA isoform produced from its genetic locus in the developing human frontal cortex and one of the highest expressed circRNAs in that system. Knockdown of this conserved circular RNA in a human neuroblastoma cell line was sufficient to induce spontaneous neuronal differentiation, measurable by increased expression of neuronal marker genes and neurite outgrowth. Depletion of circSlc45a4 in the developing mouse cortex caused a significant reduction of the basal progenitor pool and increased the expression of neurogenic regulators like Notch2, Foxp2, and Unc5b. Furthermore, a significant depletion of cells in the cortical plate after knockdown of circSlc45a4 was observed. In addition, deconvolution of the bulk RNA-seq data with the help of single cell RNA-seq data validates the depletion of basal progenitors after knockdown of circSlc45a4 in the mouse cortex and reveals an increase in Cajal-Retzius cells. Taken together, a detailed study of a conserved circular RNA that is necessary to maintain the pool of neural progenitors *in vitro* and *in vivo* is presented.

The developing mouse cortex is a good illustration for a highly spatially organized tissue and why knowledge of spatial information for each cell can be of great importance. However, obtaining transcriptome-wide and spatially resolved information from single-cells has been proven to be a challenging task. Current state-of-the-art experimental methods are either limited by the number of genes that can be detected simultaneously within a single-cell or require preexisting spatial information. Here, 3D-seq, a new experimental technique that allows unbiased, high-throughput single-cell spatial transcriptomics is introduced. 3D-seq combines a physical grid with combinatorial indexing to label single cells of any tissue in a unique way and thereby preserving the approximate spatial localization of any given cell. 3D-seq was applied to coronal slices of adult mouse brain, more than 70 cell types were identified and the 3D-seq data was used to reproduce the tissue *in silico* with single-cell resolution. Furthermore, 3D-seq is easy to adapt, can be applied to any tissue and can be combined with other technologies.

Zusammenfassung

Zirkuläre RNAs sind eine Klasse endogener, tierischer RNAs. Obwohl sie hoch abundant sind, ist weder ihre Funktion noch ihre Expression im Nervensystem bekannt. Deshalb wurde ein Katalog zirkulärer RNAs erstellt, der aus RNA-Sequenzierungen von verschiedenen Hirnregionen, primären Neuronen, Synaptoneurosomen und neuronalen Differenzierungmodellen besteht. Mit Hilfe dieser und weiterer Daten konnten tausende bisher unbekannte zirkuläre RNAs von Mensch und Maus entdeckt und analysiert werden. Zirkuläre RNAs sind außerordentlich angereichert im Säugetiergehirn, ihre Sequenz ist gut konserviert, sie sind häufig gemeinsam in Mensch und Maus exprimiert und teilweise sogar in *Drosophila* Gehirnen. Zirkuläre RNAs waren generell höher exprimiert im Verlauf der neuronalen Differenzierung, sind stark angereichert an Synapsen und oft differentiell exprimiert im Vergleich zu der korrespondierenden mRNA Isoform. Zudem korreliert die Expression von zirkulären RNAs negativ mit der Expression des RNA-editierenden Enzyms ADAR1. Induzierte Verminderung der Expression von ADAR1 führte zu erhöhter Expression von zirkulären RNAs. Somit wird ein Katalog für die Expression zirkulärer RNAs vorgestellt, der Hinweise auf die wichtigen Funktionen zirkulärer RNAs liefert.

Ausgehend von diesem Katalog wurde eine zirkuläre RNA, circSLC45A4, identifiziert. Diese zirkuläre RNA ist die Hauptisoform, die in präfrontalem, embryonalen Cortex in der 22. Schwangerschaftswoche von dem genomischen SLC45A4 Lokus exprimiert wird und generell eine der am höchsten exprimierten zirkulären RNAs in diesem System ist. Induzierte Verminderung der Expression von circSLC45A4 ist ausreichend, um die spontane neuronale Differenzierung einer humanen Neuroblastomzelllinie (SH-SY5Y) zu induzieren. Dies kann durch die verstärkte Expression neuronaler Markergene und das verstärkte Neuritwachstum belegt werden. Verminderung der Expression von circSLC45A4 im sich entwickelnden, embryonalen Mauscortex verursacht eine signifikante Reduktion von basalen Progenitoren während neurogene Regulatoren wie Notch2, Foxp2 und Unc5b verstärkt exprimiert werden. Außerdem wurde eine signifikante Reduktion von Zellen in der kortikalen Platte nach Depletion von circSLC45A4 gemessen. Weiterhin konnten die im manipulierten Mauscortex gesammelten RNA-Sequenzierergebnisse mit Hilfe von Daten einer Einzelzellsequenzierung dekonvoliert werden. Dies bestätigte die Reduktion der basalen Progenitoren und zeigte zusätzlich, dass vermehrt Cajal-Retzius Zellen auftraten. Somit wird eine detaillierte Untersuchung einer konservierten zirkulären RNA, die notwendig für den Erhalt von neuronalen Progenitoren ist, präsentiert.

Der sich entwickelnde Mauscortex ist ein gutes Beispiel für ein Gewebe, das eine starke räumliche Organisation aufweist und verdeutlicht damit, warum die Kenntnis der räumlichen Lage für jede einzelne Zelle von großer Wichtigkeit sein kann. Dennoch ist es weiterhin eine große Herausforderung gleichzeitig Transkriptom-weite and räumlich aufgelöste Sequenzierdaten im selben Experiment zu sammeln. Bisher existierte keine Methode die beides leisten konnte. Die vielversprechendsten Methoden sind entweder begrenzt in der Anzahl an Transkripten, die gleichzeitig in einer Einzelzelle gemessen werden können, oder benötigen bereits anderweitig existierende, räumliche Informationen. Hier wird eine neue

experimentelle Methode vorgestellt, die Hochdurchsatz RNA-Sequenzierung von Einzelzellen mit räumlicher Auflösung zulässt, ohne vorherige Kenntnisse des Systems zu benötigen: 3D-seq. 3D-seq vereint die Applikation eines physischen Gitters am Gewebe mit kombinatorischem Indizieren, so dass Einzelzellen individuell und räumlich markiert werden können. 3D-seq wurde an koronalen Schnitten von adultem Mausgehirn etabliert. Hierbei konnten mehr als 70 Zelltypen identifiziert werden und die so erhaltenen Daten wurden zur Reproduktion des Gewebes *in silico* genutzt. Zudem ist 3D-seq ein leicht zu adaptierendes Protokoll, das an jedem Gewebe angewendet werden kann und ebenfalls leicht mit anderen Techniken kombiniert werden kann.

Acknowledgments

First of all I would like to thank my supervisor and mentor *Prof. Dr. Nikolaus Rajewsky* for showing me what it means to be a scientist, his support, and thorough discussions in the past 5 years. I am grateful to him for all the doors he opened and opportunities he gave me to develop my knowledge and abilities. Last but not least I have enjoyed his trust in me when I came to him with new project ideas, when he gave me challenging tasks or encouraged me to give talks in front of large audiences.

Furthermore, I would like to thank a number of people that have enriched my PhD and supported me in this exciting time:

Sebastian Kadener for being an enthusiastic and motivating co-mentor and the members from his lab, especially *Nagarjuna Pamadurti, Osnat Bartok, Mor Hanan*, and *Reut Ashwal-Fluss* for fruitful collaborations, discussions, and their friendship.

Markus Landthaler and *Marina Chekulaeva* for regularly and thoroughly discussing my projects.

The *Boehringer Ingelheim Fonds* for financial support during 3 years of my PhD, offering contacts to fellow scholarship holders and challenging me with absurd hikes.

Signgene for having me participate in exciting scientific symposia and the opportunity to get to know great people and Israel so well.

Margareta Herzog and Alexandra Tschernycheff for everything they are taking care of in our lab. Without them it would fall apart.

Agnieszka Rybak-Wolf and *Petar Glazar* for being a great team when we published together.

Gwendolin Matz, *Salah Ayoub*, *Anastasyia Boltengagen*, and *Luisa Schreyer* for all their support with experiments and various tasks in the lab.

Nikolaos Karaiskos for being my partner in crime in the 3D-seq project, overcoming several setbacks as a team and even convincing people to give us money for this project.

Koshi Imami and *Henrik Zauber* from *Matthias Selbach*'s lab for MS measurements either of the circRNA pulldown or the potential circRNA translation, respectively.

Daniel Cavalli, *Martina Dori*, *Simone Massalini*, and *Federico Calegari* for *in vivo* experiments in the circSLC45A4 project and the discussion of their results.

Filippos Klironomos and *Andrei Filipchyk* for help with miRNA analysis.

Panagiotis Papavasileiou, Tamas Sztanka-Toth, Umut Kilik, Benedikt Obermayer, Angelika Kusch, Dennis Gürgen, Druska Dragun, Annett Böddrich, Erich Wanker, Steffen Fuchs, Johannes Schulte, Angelika Eggert, Erika Urdaneta, Benedikt Beckmann and many more people for collaborating with me on several projects that have not found their way into this doctoral thesis but were not least important to me.

And finally, I would like to thank my *family and friends* for their support and encouragement at all times during my PhD. With greatest sincerity I want to thank my *husband*, *Benjamin Sünkel*, especially, all the hours he has taken to discuss my experiments, listening to endless yackety-yak about puzzling results, enjoying a Valentine's day dinner despite having to listen to new project ideas, showing up in the lab with late night snacks or keeping me company when experiments were running until after midnight. But despite all of this making sure that I don't forget that there is a life outside the lab and taking on all the great adventures of life with me.

Author contributions

All projects have been guided, discussed, and supervised thoroughly by Prof. Dr. Nikolaus Rajewsky.

Related to results chapter 7.1 - Circular RNAs in the Mammalian brain are highly abundant, conserved and dynamically expressed:

Large parts of this chapter have been published in Molecular Cell (see 12 - Publication list). All results obtained within this project are the result of a highly collaborative work between Dr. Agnieszka Rybak-Wolf, Petar Glazar and the PhD candidate. A. R.-W., P.G. and C.S. designed all experiments and analysis thereof. Experiments in mouse were conducted by A. R.-W., experiments in human were conducted by C.S. P.G. wrote the code for the majority of computational analysis, except for Figure 11 – C + F (C.S.), Figure 12 – E + F (C.S.), Figure 13 – C (Marvin Jens), Figure 14 – A + G (Andranik Ivanov and Reut Ashwal-Fluss (Hebrew University, Jerusalem, Israel). Synaptoneurosome samples were provided by Marie Öhman (Stockholm University, Stockholm, Sweden).

Related to results chapter 7.2 - CircSLC45A4 is required to keep neuronal cells in a progenitor state in cell culture systems and in the mammalian brain:

Large parts of this chapter have been submitted to Cell Reports (in revision) (see 12 - Publication list). C.S. designed and conducted all experiments and analysis, including computational analysis of high-throughput data. *In utero* electroporation, staining, quantification, and sorting of GFP+ cells from mouse cortices were done by Daniel Cavalli and Simone Massalini in Federico Calegari's laboratory (CRTD, Dresden).

Related to results chapter 7.3 - 3D-seq: a method to integrate single-cell and spatial transcriptomics:

C.S. proposed the idea & project, and developed a detailed plan together with Nikolaus Rajewsky and Nikolaos Karaiskos. C.S. designed and conducted all experiments. N.K. wrote all code for computational analysis. C.S., N.K. and N.R. discussed all results and necessary analysis.

1 Table of Contents

2 Abbreviations

AP	Apical progenitor
ATRA	All-trans retinoic acid
BP	Basal progenitor
bp	base pair
BSA	bovine serum albumin
Bst	brainstem
cb	Cerebellum
CDS	coding sequence
ch	Cerebral hemisphere
circRNA	circular RNA
CNS	central nervous system
CP	Cortical plate
ddH_2O	double distilled water
dTh	Dorsal thalamus
E	Embryonic day
EB	Embryoid body
EMCV	Encephalomyocarditis virus
FAM	6-Carboxyflurescein
HA	Hyperpallium apicale
HC	hippocampus
hnRNA	heterogenous nuclear RNA
IHA	Interstitial nucleus
IP	intermediate progenitor, immunoprecipitation; context dependent
IRES	Internal ribosome entry site
IT	Intratelencephalic neuron
IUE	*In utero* electroporation
IVT	*in vitro* transcription
IZ	Intermediary zone
KD	knockdown
lncRNA	long non-coding RNA
Lx	Layer x
m	Medulla oblongata
Md	Dorsal mesopallium
miRNA	micro RNA
MMLV	Moloney murine leukemia virus
MZ	Marginal zone
ncRNA	Non-coding RNA
NEC	Neuroepithelial cell
NEU	Neuron
ob	Olfactory bulb
ot	Optic tectum
PA	Polyamide
PC	Polycarbonate
PC	Polycarbonate
PDMS	polydimethylsiloxane
PFA	paraformaldehyde
pg	Pituitary gland
PLA	polyamide
PLL	poly-L-lysine
PP	preplate
qPCR	Quantitative real time PCR
RBP	RNA-binding protein
RCM	reverse complementary sequence matches
RGC	radial glial cells
RNAi	RNA interference
RNA-seq	RNA sequencing
RT	**Reverse transcription**
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SLC	SLC45A4
Slc	Slc45a4
SVZ	Subventricular zone
TEMED	Tetramethylethylenediamine
TPM	transcripts per million
UNG	uracil-N-glycosylase
UTR	Untranslated region
VZ	Ventricular zone
WG	Gestational week

3 Introduction

3.1 Circular RNAs are a new class of functional RNAs

3.1.1 Emergence and biogenesis of circular RNAs

In 1957, Francis Crick postulated in a lecture at the University College London 'the central dogma of molecular biology' (Crick, 1970, 1958). He proposed the idea of information flow, meaning the transfer of information, from DNA to RNA to protein and certain variations thereof (Figure 1 - A). This proposal included a side note on the likely existence of RNA molecules that function as adapter molecules, which were indeed shortly thereafter identified as tRNAs. Although, this conceptualization of information flow in a cell is powerful, it is a simplified concept to describe the complex regulatory networks found in each cell. The view on how complex the regulatory networks are, was considerably broadened by discoveries that were made in the field of RNA biology (Morris and Mattick, 2014). The discovery of heterogeneous nuclear RNAs (hnRNAs) in 1969 (Warner et al., 1966) led to speculations about the existence of a broad RNA-based regulatory network (Britten and Davidson, 1969). The existence of a broad RNA-based regulatory network was further spurred by the discovery of introns in 1977 (Berget et al., 1977; Chow et al., 1977), of ribozymes (Guerrier-Takada et al., 1983; Kruger et al., 1982), of the regulatory function of miRNAs (Fabian et al., 2010; Krek et al., 2005; Lagos-Quintana et al., 2001; Lee et al., 1993; Reinhart et al., 2000), of RNA interference (Elbashir et al., 2001a; Fire et al., 1998), and of long noncoding RNAs (lncRNAs) (Guttman et al., 2009; Mercer et al., 2009).

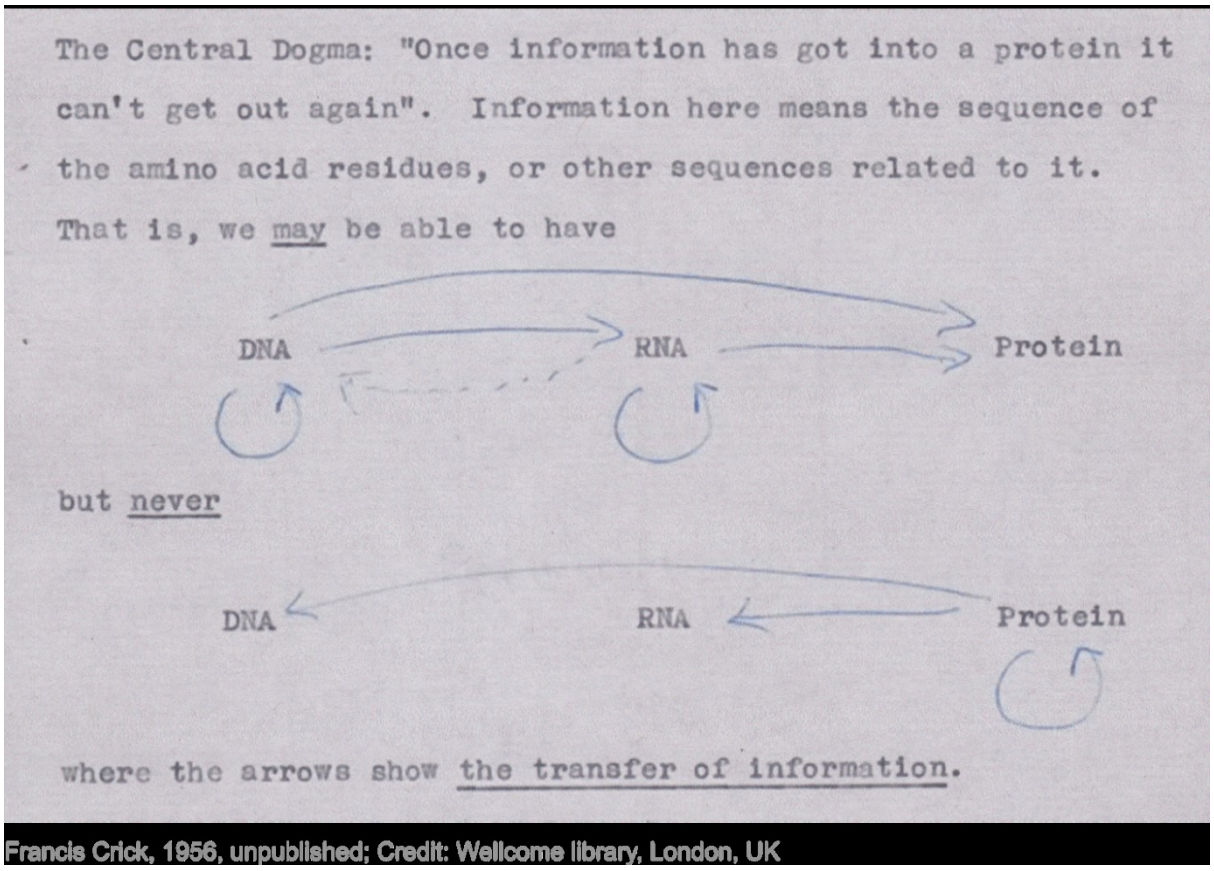


Figure 1: Central Dogma of Molecular Biology (Francis Crick, 1956)
Unpublished note by Francis Crick depicting his postulated 'central dogma of molecular biology'.

The various already mentioned RNA species have in common that all of them have a free 3'- and 5'-end. However, circular RNAs (circRNAs) are covalently closed RNA molecules and have been discovered several decades ago. They came back into focus with the broad description of them as a whole class of RNA molecules (Memczak et al., 2013). Already in the 1970s, certain plant pathogens, viroids, were described to consist of single-stranded, covalently closed circular RNA molecules (Sanger et al., 1976). Shortly after this discovery, circular RNA structures were discovered in cytoplasmic RNA extracts from HeLa cells (Hsu and Coca-Prados, 1979). The authors speculated that these circular RNAs are indeed covalently closed, as not even highly denaturing conditions are able to disrupt their circular structure. Furthermore, they suggested that circular RNAs could be remains from a viral infection. But if they were produced by HeLa cells themselves, circRNAs could be serving as efficient translation templates. And indeed, the genome of the animal Hepatitis delta virus was also shown to be a covalently closed circular RNA that could hypothetically be a template for translation (Kos et al., 1986). It was also speculated that due to their small size and circularity, circRNAs had an enhanced probability of survival in error-prone, primitive self-replicating RNA systems and that complete replication could occur without the need for initiation or termination signals (Diener, 1989). Hence, it was indicated that circular RNAs could be remnants of the RNA world.

However, circular RNAs are not only parts of viruses or remnants of the RNA world. Instead, they were shown to be produced from eukaryotic genes. The gene DCC for example produces at least 4 different circular RNAs that although lowly expressed were detectable by a PCR assay (Nigro et al., 1991). Another example is the human ETS-1 gene, which was shown to be non-polyadenylated, lowly expressed and not produced from genomic rearrangements or pseudogene transcription but instead from a splicing reaction that yields circular molecules (Cocquerelle et al., 1992, 1993). The authors also highlight, that the circularly spliced exons of ETS-1 are surrounded by unusually large introns just like the DCC circRNAs. Also, they show that circETS-1 is cytoplasmic and unusually stable, as it is still detectable after 48 h of transcriptional block with actinomycin D. They considered circular RNAs a result of missplicing that could potentially help shed light on splicing reactions in general. Interestingly, it became clear that circular RNAs are not always lowly expressed: a circular RNA originating from Sry was determined to be the highest expressed transcript isoform from this gene in testis (Capel et al., 1993). The authors reported that this circular RNA was cytoplasmic and not substantially bound to polysomes although it would contain a 393 amino acid long open reading frame. However, the mechanism of this circular RNA remained elusive. *In vitro* experiments in nuclear HeLa extracts indicated that indeed the splicing machinery is responsible for joining exons in an inverted order and thereby producing circular RNAs (Braun et al., 1996; Pasman et al., 1996), a process that is also known as backsplicing (Figure 2 - A).

Noteworthy, recent transcriptome wide studies revealed that circRNAs are an abundant phenomenon and not rare missplicing events. Indeed, hundreds of circRNAs were identified in Archaea (Danan et al., 2012), human cell lines (Jeck et al., 2013; Memczak et al., 2013; Salzman et al., 2012), mouse and nematodes (Memczak et al., 2013), and *Drosophila* (Ashwal-Fluss et al., 2014; Westholm et al., 2014). Many of these circRNAs are highly abundant, tissue

and developmental-stage specific (Memczak et al., 2013). Most circRNAs are indeed produced from exons, however circRNAs originating from all types of genomic regions have been identified– introns, 5'UTRs, 3'UTRs, intergenic regions etc. (Memczak et al., 2013). Furthermore, exons containing circRNAs were more conserved in the third codon position, indicating evolutionary constraints on the nucleotide level in addition to selection at the protein level (Memczak et al., 2013). An important tool to identify circular RNAs is RNase R, a 3'-5' exonuclease from *Escherichia coli* that digests linear RNA and Y-structure RNA efficiently, while circular RNAs and lariat RNA are left intact (Suzuki et al., 2006). To discriminate between circular RNAs and intron lariats a debranching assay, Sanger sequencing of the head-to-tail splice junction, nicking assays or a head-to-tail junction specific Northern blot can be used (Memczak et al., 2013).

Converging evidence exists that circRNAs are processed *co*-transcriptionally. Backsplicing seems to be mediated by the spliceosome, as circRNAs are detected in chromatin-associated RNA and drugs arresting spliceosome activity significantly reduce circRNA production rates (Ashwal-Fluss et al., 2014; Starke et al., 2015; Zhang et al., 2014). In addition, the rate of circRNA production is highly affected by the presence of canonical splice sites in surrounding exons and correlates positively with the elongation rate of RNA Polymerase II (Zhang et al., 2016). An RNAi screen in *Drosophila melanogaster* identified several pre-mRNA processing factors that modulate circular RNA levels. Surprisingly, circular RNA expression increased when core spliceosomal factors were limited (Liang et al., 2017). Backsplicing is enhanced, when the circRNA-surrounding introns possess highly complementary sequences. These highly complementary sequences are often transposable elements, like *Alu* elements (Ivanov et al., 2015; Liang and Wilusz, 2014; Zhang et al., 2014). Accordingly, circRNAs are often surrounded by introns that contain *Alu* elements (Jeck et al., 2013). Most splicing in human, mouse, and *Drosophila* occurs co-transcriptionally in the nucleus (Brugiolo et al., 2013). CircRNAs are pre-dominantly localized in the cytoplasm (F. Torti, personal communication, 2013). Hence, they have to be exported from the nucleus. Their nuclear export is length dependent and requires the DexD-Box helicases DDX39A or DDX39B, at least in a human cell culture model (Huang et al., 2018). circRNAs are extremely stable, likely due to the lack of free 3'- or 5'-ends (Enuka et al., 2016; Hansen et al., 2013; Jeck et al., 2013; Memczak et al., 2013; Salzman et al., 2013; Zhang et al., 2016).

Additionally, several RNA-binding proteins were shown to modulate circRNA biogenesis, including the splicing factor *muscleblind* (MBL/MBNL1), and the RNA-binding proteins QKI, ADAR1 and ILF3/NF90/NF110 (Figure 2 – B) (Ashwal-Fluss et al., 2014; Conn et al., 2015; Ivanov et al., 2015; Li et al., 2017). MBL/MBNL1 protein binds the flanking introns of its own transcript and promotes circularization (Ashwal-Fluss et al., 2014). That this circularization can occur, has been described already in 2006 (Houseley et al., 2006), however, its function remained unclear. Modulation of MBL amounts strongly affects the formation of circMBL and of at least three other circRNAs. Since backsplicing events compete with mRNA splicing, the generation of circRNAs will come at the expense of its linear transcript. This suggests that circRNA biogenesis by itself may regulate gene expression, e.g. by preventing the production of the mRNA (Ashwal-Fluss et al., 2014). QKI is modulating the biogenesis of a subset of circRNAs during mesenchymal to epithelial transition by enhancing the pairing of circRNA

surrounding introns (Conn et al., 2015). The RNA editing enzyme ADAR1 was also recently shown to modulate the generation of circRNAs, likely by editing and disrupting inter- or intra-intronic complementary sequences (Ivanov et al., 2015; Rybak-Wolf et al., 2015). DHX9 is a nuclear helicase that binds Alu elements and modulates RNA processing. Silencing DHX9 causes an increase in host genes that encode circRNAs in addition to a general increase in circRNA expression (Aktaş et al., 2017). In addition, NF90/NF110 act as trans-factors that promote circularization in the nucleus by stabilizing the transient intronic RNA pairs and silencing of these genes reduced nascent circRNA production (Li et al., 2017). These proteins are usually exported to the cytoplasm upon activation of the immune response after viral infection. This reduces their ability to bind nascent transcripts in the nucleus and suppresses the production of a set of circRNAs.

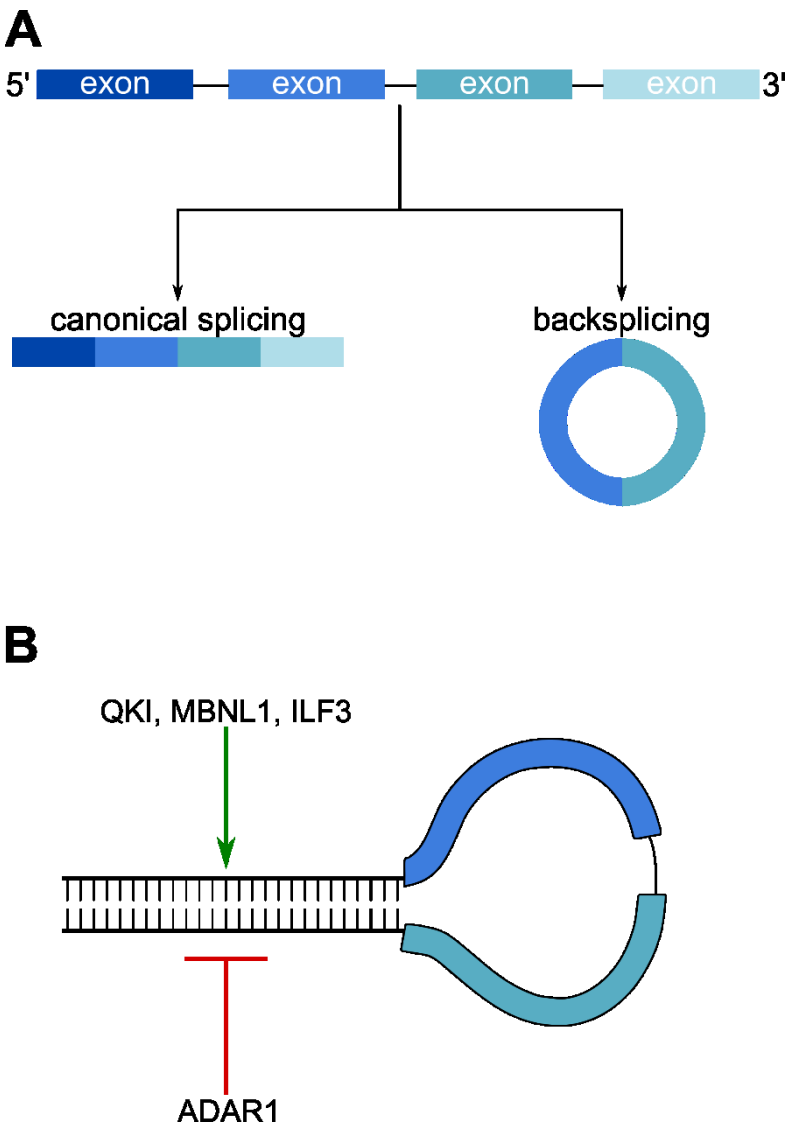


Figure 2: Circular RNAs are produced by backsplicing.
(A) – Circular RNAs are produced by backsplicing, were the 5' end and the 3' end of one or several exons are joined in an inverted order. mRNAs on the other hand are spliced linearly. **(B)** – QKI, MBNL and ILF3 promote circular RNA production, while ADAR1 interferes with their biogenesis.

In general, neural tissues express wider repertoires of alternatively processed RNA isoforms (Wang et al., 2008; Yeo et al., 2004). Therefore it is not surprising that circRNAs are particularly enriched in the brain (Rybak-Wolf et al., 2015; Venø et al., 2015; You et al., 2015). Interestingly, many circRNAs are expressed in specific brain regions and are even enriched in axons and synapses. Furthermore, a subset of circRNAs are strongly modulated upon neuronal differentiation of *in vitro* cell culture systems (Rybak-Wolf et al., 2015; You et al., 2015) as well as during specific stages of porcine brain development (Venø *et al*, 2015). Indeed, circRNAs might be involved in brain diseases, as many of them accumulate with age in the CNS of *Drosophila melanogaster* (Westholm et al., 2014). Additionally, it was recently reported that circRNAs with at least moderate expression in the brain are highly conserved between mouse and human (Rybak-Wolf et al., 2015). More specifically, in this subgroup, 60% of mouse circRNAs that are expressed in humans share the exact splice sites. Another 20% of circRNAs share at least one conserved splice site between mouse and human. Moreover exonic sequences giving rise to circRNA molecules display a higher occurrence of conserved 7mers than exonic sequences giving rise to only mRNAs (Rybak-Wolf *et al*, 2015).

3.1.2 Functions of circular RNAs

A number of potential circRNA functions have been proposed (Hentze and Preiss, 2013; Memczak et al., 2013): as microRNA or RNA-binding protein (RBP) delivery vehicle, as RBP sponge or assembly factory for RBP factories, as regulators of RBP function or (m)RNA expression, or as template for translation itself (Figure 3 - A).

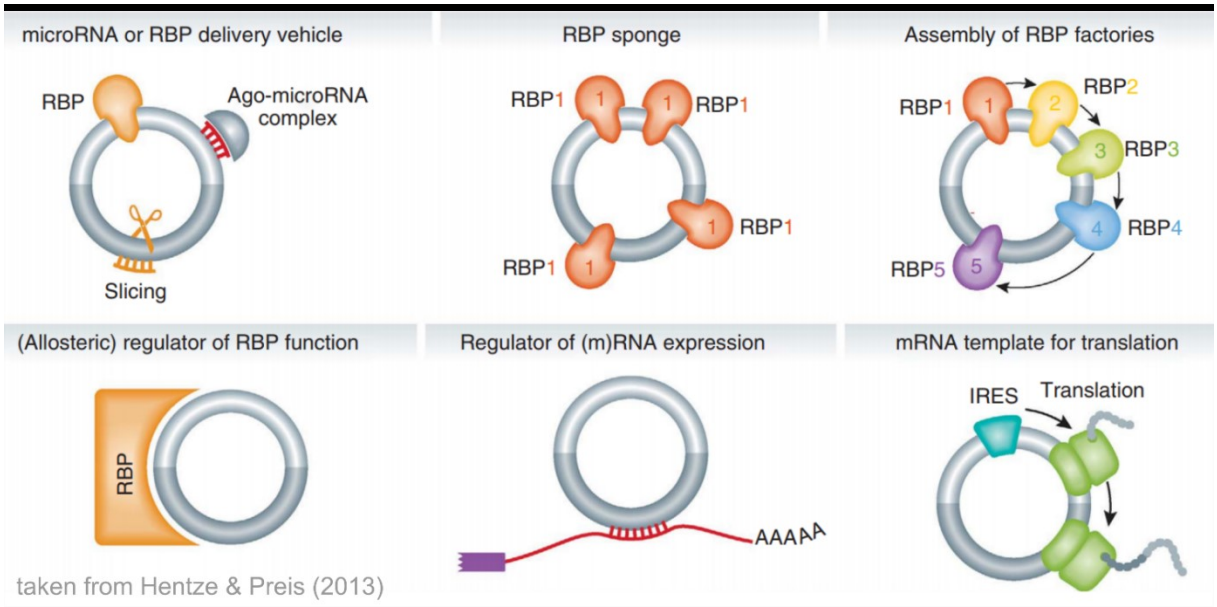


Figure 3: Potential functions of circular RNAs (Hentze and Preiss, 2013).
There is no evidence for a general function of circRNAs. However (Hentze and Preiss, 2013) propose that circRNAs could act as delivery vehicles, RNA-binding protein (RBP) sponges, help assembling RBP factories, regulate RBP function, regulate expression or even serve themselves as templates for translation.

So far, few examples of the function of circular RNAs exist. The known examples for circRNA function are diverse and indicate that circular RNAs as a class share their topology but not necessarily their mode of function. circFOXO3 for example functions as protein hub and helps

assemble a larger protein complex. It was shown to be essential in forming a ternary complex with p21 and CDK2 and by this regulates cell cycle progression (Du et al., 2016).

Translation was suggested as one likely mode of action for some of the first identified circular RNAs (Hsu and Coca-Prados, 1979; Kos et al., 1986). Since then it has been debated whether circRNA translation is a common phenomenon or whether it occurs at all. circSry, as prominent circRNA example, was not found at polysomes (Capel et al., 1993) and it was speculated that the circular form of Sry is specifically made to avoid translation. Then, *in vitro* experiments were conducted that showed how inefficiently circular templates were bound to eukaryotic (wheat-germ) (Kozak, 1979). While an mRNA with the same sequence was bound efficiently, despite its lack of a polyA-tail and a m^7G-cap. Thus, it seemed that the circular structure of the RNA template prevented its efficient binding to the ribosome. Interestingly, circular RNA templates were efficiently bound to prokaryotic ribosomes, even more so than linear RNAs with the same sequence. This finding was reproduced for other circular and linear messages (Konarska et al., 1981). Both of these publications have been challenged by the discovery of internal ribosome entry sites (IRES). A circular RNA containing a viral IRES from encephalomyocarditis virus (EMCV) was translated *in vitro* in rabbit reticulocyte lysate, even though it was less efficiently translated than a linear template with the same sequence (Chen and Sarnow, 1995). Later on it was also demonstrated that a circular GFP containing an EMCV IRES can be translated in human HEK 293 FlpIn T-Rex cells (Meyer et al., 2015; Wang and Wang, 2015). However, these are *in vitro* experiments with sequences that would not occur naturally. Hence, it is even more exciting that examples of translated, endogenous circular RNAs exist: circZNF609 was identified in an expression profiling screen of *in vitro* muscle differentiation in mouse and human to be regulating myoblast proliferation. Legnini et al. found an open reading frame running over the head-to-tail junction, showed that circZNF609 is associated with the heavy polysome fraction, identified a translation product of the expected size with a specific antibody and showed that the translation of this circORF was splice-dependent and cap-independent (Legnini et al., 2017). Additionally, to this example of a translated circRNA in mouse and human, an example of an endogenous translated RNA in *Drosophila melanogaster* exists. Here, ribosome foot printing on *Drosophila melanogaster* samples revealed ribosome-associated circular RNAs. Many of the ribosome-associated circRNAs had evolutionary conserved termination codons that were also matched by ribosome foot printing reads. Furthermore, it was shown that a circMbl-derived protein is detected by mass spectrometry and that circMbl and several other circRNAs are translated *in vitro* and *in vivo* in a cap-independent manner (Pamudurti et al., 2017). Stagsted et al. classified a subset of circRNAs as AUG circRNAs. Those are circRNAs that include the canonical translation start site that is also used in the corresponding mRNA. AUG circRNAs are conserved and are more abundantly expressed in comparison to other circRNAs. They are circularized in an *Alu*-independent manner and are not found to act as template for translation based on RiboSeq analysis. Stagsted and colleagues challenge the notion that many circRNAs can be a template for peptides as their analysis demonstrates that most, but not all, circRNAs are not bound to ribosomes (Stagsted et al., 2018). In summary, translation of circRNAs seems to occur only under certain circumstances, maybe in stress conditions, or only in specialized cell compartments or at specific developmental steps. Cap-independent translation in eukaryotes is a specialized process as well that is mostly observed in stress, infection, and disease

(Merrick, 2004).Our understanding of circRNA translation will broaden as we understand the involvement of potential translation regulators, like the exon-junction complex (Le Hir et al., 2016), and RNA modifications, like N6-methyladenosine (Meyer et al., 2015; Wang et al., 2015; Yang et al., 2017).

CDR1as is a central part of a regulatory network of non-coding RNAs (Kleaveland et al., 2018; Piwecka et al., 2017). Cdr1as is massively bound by miR-7, in > 120 positions in mouse, and miR-671 (Hansen et al., 2013; Memczak et al., 2013). Depletion of the entire locus in a knockout mouse model revealed that loss of Cdr1as impairs sensorimotor gating, a phenotype often observed in neuropsychiatric disorders, and causes dysfunctional synaptic transmission. Molecularly, knockout of the Cdr1as locus specifically deregulates miR-7 and miR-671 levels in several brain regions by post-transcriptional effects. This in turn may cause induction of immediate early genes, such as Fos, a known miR-7 target. Kleaveland et al. expands the view on this ncRNA network by a long-noncoding RNA, Cyrano that was shown to destruct miR-7. miR-7 on the other hand can repress Cdr1as expression potentially by promoting the slicing of Cdr1as by miR-671, while Cdr1as itself stabilizes miR-7, as was shown before by Piwecka et al.

Chen et al. suggest that circRNAs are used as sensors for self/non-self discrimination. Exogenous circRNAs are triggering an innate immune response via RIG-1, that protects cells against viral infection (Chen et al., 2017). It was also shown that the self/non-self discrimination depends on the human nature of the spliced-out intron.

Besides these examples of circRNA function, many other modes of action could exist for this versatile class of RNA molecules.

3.2 Cortical development in Mammalia

The brain is a highly complex organ that is composed of multiple tissues. The brain and spinal cord form the central nervous system. The vertebrate brain consists of an olfactory bulb, a cerebellum, cerebral hemispheres, a medulla oblongata, an optic tectum and a pituitary gland (Northcutt, 2002). All of these regions exert specialized functions (Table 1) and have changed their sizes and shapes over 500 million years of vertebrate evolution (Figure 4 - A). Especially the cerebral hemispheres have grown tremendously from lamprey to mammals.

Table 1: Vertebrate brain regions and their function (Purves et al., 2004)

Brain region	Function
Olfactory bulb	Relaying olfactory signals from cranial nerve to higher centers
Cerebellum	Hindbrain structure coordinating motor functions, balance and posture
Cerebral Hemispheres	Processing of visual, auditory, somatosensory, and olfactory information. Directing conscious motor functions, speech, language, memory and learning.
Medulla Oblongata	Part of the brain stem, contains autonomic functions (cardiac, vomiting, vasomotor, and respiratory centers)
Optic Tectum	Central region to processing visual signals in many vertebrates (in mammals together with visual cortex)
Pituitary Gland	Hormone and neuropeptide secreting region, controlling e.g. blood pressure, growth, thyroid glands, and metabolism

The cerebral neocortex is located within the cerebral hemispheres and a key feature of mammalian brain evolution (Striedter, 2005). The mammalian cerebral neocortex is six-layered (Figure 4 - B), and plays an important role in higher cognitive, emotional, sensory and motor functions. It has expanded tremendously compared to lower vertebrates (Kwan et al., 2012) and this expansion of the neocortex is thought to be highlighted by the gyrification of the brain found in certain mammals, e.g. human and other primates. However, this feature is almost absent from rodents, except for ferrets (Striedter, 2005). Although the six-layered structure of the mammalian neocortex is not found in other vertebrates, the general signal-relaying cell types of the pallium (layers of white and grey matter of the upper cerebrum) are found in all vertebrates (Briscoe and Ragsdale, 2018). It is thought that already the last common ancestor had neurons receiving input signals from the dorsal thalamus (dTh), relaying them first to intratelencephalic (IT) neurons and then to output neurons before the signal is passed on to the brain stem (Bst). Many different processes, e.g. neuronal migration, neuronal differentiation, dendro-/axo-/synaptogenesis and gliogenesis, have to be coordinated to ensure proper development of the mammalian neocortex during development (Komuro and Rakic, 1998; Kriegstein and Noctor, 2004; Kwan et al., 2012; Taverna et al., 2014).

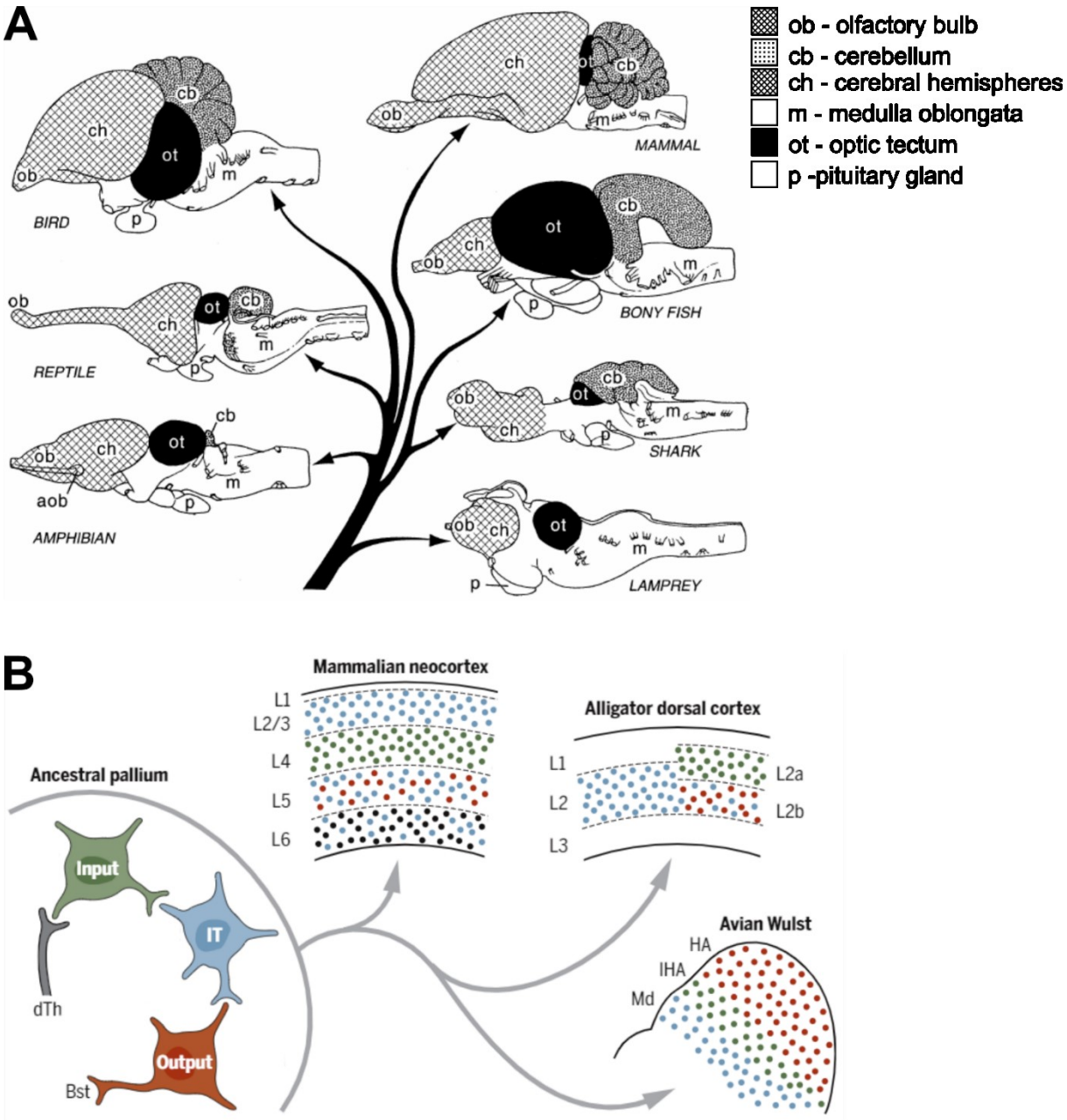


Figure 4: The six-layered neocortex is a unique feature of mammalian evolution.
(A) – Vertebrates share the gross anatomical structure of their brains. While the overall function of these regions remained similar over 500 million years of evolution, size and shape of the brain regions changed. ob – olfactory bulb; cb – cerebellum; ch – cerebral hemispheres; m – medulla oblongata; ot – optic tectum; p – pituitary gland. Image taken from (Northcutt, 2002). **(B)** – The main cell types of the vertebrate pallium (layers of white and grey matter of the upper cerebrum) are comparable between avians, non-bird reptiles (alligator) and mammalia. The common ancestor is thought to have had input (green), output (red) and intratelencephalic (IT, blue) pallial neurons that receive information from the dorsal thalamus (dTh) and relay them to the brainstem (Bst). The cytoarchitecture changed during evolution, the general cell types are shared. Lx – layer x; HA – hyperpallium apicale; IHA – interstitial nucleus of the hyperpallium apicale; Md – dorsal mesopallium. Image taken from (Briscoe and Ragsdale, 2018).

Most of our knowledge about neocortical development was obtained from studies of mouse. In general, cortical projection neurons are born in the dorsal germinal zones of the telencephalon, the ventricular and subventricular zones (VZ and SVZ) (Angevine and Sidman, 1961). They radially migrate to their final destination in the upper layers of the developing cortex, either guided by the processes of radial glial cells (= locomotion) or independent of radial glial cells (=translocation) (Gupta et al., 2002). Early in mouse development, before embryonic day 10 (E10), only the ventricular zone and the marginal zone (MZ) exist (Figure 5 - A). Until this point the neural progenitor pool is expanded. At around E10.5, the first wave of

cortical neurons is generated to form the preplate (PP). At E11.5 another wave of cortical projection neurons is leaving the VZ to form the nascent cortical plate (CP) thereby splitting the PP into marginal zone (MZ) and subplate (SP). As embryonic development continues, further waves of cortical neurons are leaving the VZ and SVZ to form the upper layers of the cortex (L6-L1). The cortex is formed in an inside-first, outside-later fashion. This means, that early born neurons are populating the lower/deeper layers of the cortex (L6), while later born neurons migrate past them to form the upper/superficial layers (L1). However, many different processes have to be spatiotemporally coordinated to allow successful generation of neurons from neural stem and progenitor cells (Götz and Huttner, 2005; Mora-Bermúdez et al., 2013; Noctor et al., 2007; Sun and Hevner, 2014). Yet, any misregulation of one of these processes can prevent proper development of the cortex.

Early in vertebrate development, neural fate is induced in the ectoderm and the neural plate is formed. Neurulation then forms the neural tube from the neural plate. The wall of the neural tube is made up from neuroepithelial cells (NEC) (Figure 5 - B) (Götz and Huttner, 2005; Paridaen and Huttner, 2014; Taverna et al., 2014), which move their nuclei according to the cell cycle phase they are in: nuclei go to the ventricular surface to undergo mitosis. Then, at the onset of neurogenesis around E10.5 the NECs switch identity and become radial glial cells (RGC). RGCs give rise, directly or indirectly, to all neurons and glial cells during development. NECs and RGCs are often referred to as apical progenitors (AP) as they exhibit apico-basal polarity. These apical and basal processes expand over the entire neuroepithelium. However, when NECs switch their identity to RGCs they lose tight junctions but keep adherens junctions to connect themselves to the apical surface of the neuroepithelium. Furthermore, they induce astroglial marker genes like GLAST and BLBP. The pathways behind this transition are poorly understood. However, it was shown that transcription factors of the Hes family are involved (Hatakeyama et al., 2004) as well as transient expression of Fgf10 (Sahara and O'Leary, 2009). With the switch to RGCs the cell divisions become asymmetric, each RGC giving rise to one daughter RGC and one differentiating cell. In the mammalian neocortex those differentiating cells can either form intermediate progenitors (IP) that are a type of basal progenitor (BP) only giving rise to neurons. Additionally, BPs (referred to as bRG in Figure 5) with glial characteristics that are able to self-renew can be formed from RGC division. These newborn neurogenic daughter cells, IPs and BPs, lose their apical endfoot that connects them to the apical surface in order to migrate basally. This loss is regulated by proneural genes that either induce downregulation of cadherins that mainly maintain this attachment through adherens junctions or by inducing the abscission of the apical endfoot altogether. IPs will give rise to two newborn neurons, BPs (=bRGs) to another BP and a neuron. Those newborn neurons migrate in waves to the pial surface of the brain across the intermediate zone and settle down in the cortical plate, where they mature and form synaptic connections between each other (Angevine and Sidman, 1961; Götz and Huttner, 2005; Gupta et al., 2002; Paridaen and Huttner, 2014; Taverna et al., 2014).

Additionally, post-mitotic neurons in the upper cortical layers can influence the progenitor cells in this system (Paridaen and Huttner, 2014). It was shown that Cajal-Retzius cells, a neuron type born very early in the neocortex, dynamically regulates apical progenitor maintenance and the switch to neurogenesis and gliogenesis in the neocortex (Lakomá et al., 2011).

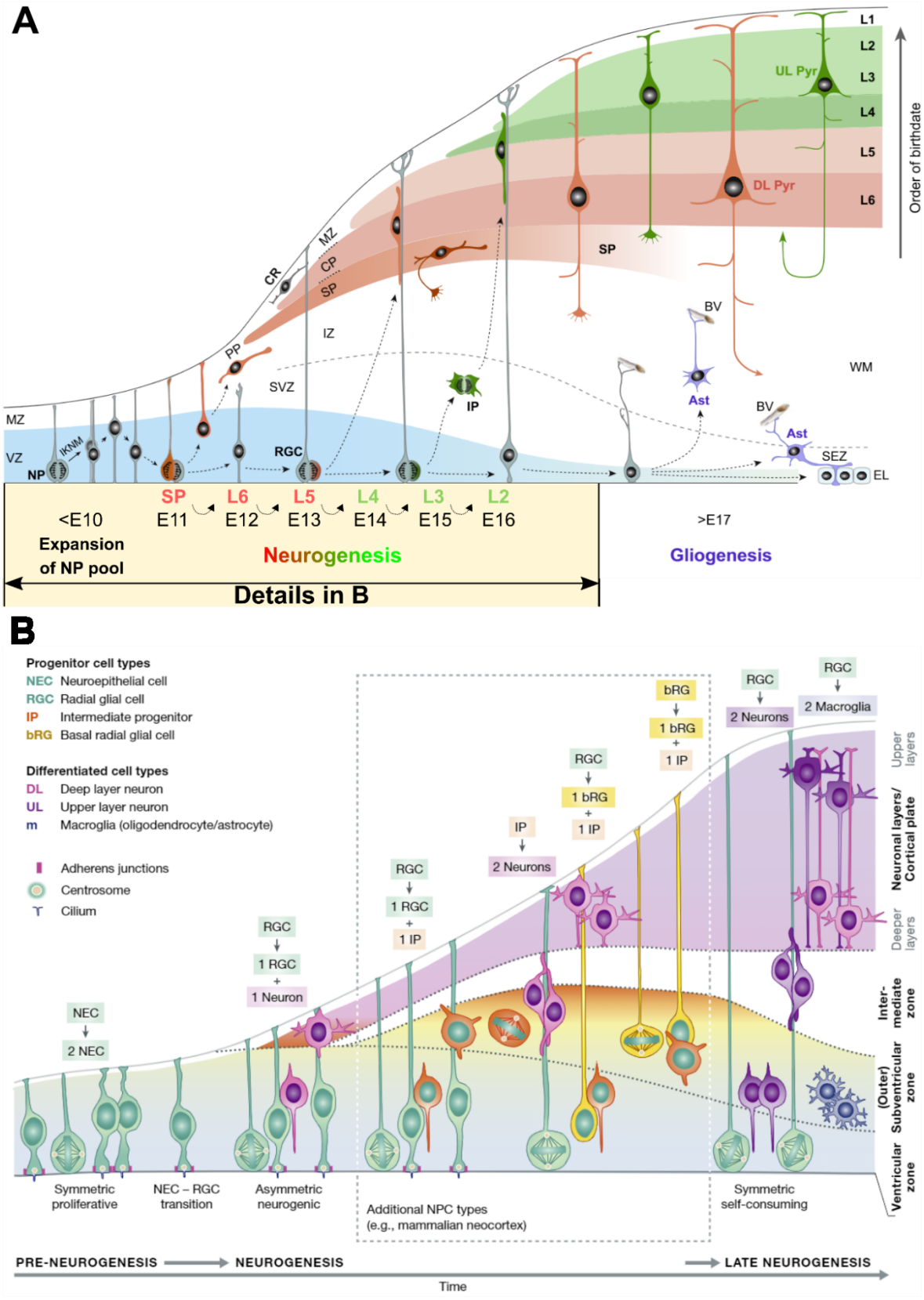


Figure 5: Development of the six-layered mouse cortex.
(A) – Before the onset of neurogenesis neural progenitors (NP) in the ventricular zone (VZ) increase their pool by cell division, in which they undergo interkinetic nuclear migration (IKNM). At ~E11.5 NPs change identity to radial glial cells (RGC) and undergo asymmetric neurogenic divisions. The first neurons build the preplate (PP) that is then split by further neurons in the marginal zone (MZ) and the subplate (SP). Neurons settling in the cortical plate (CP) will later form the cortical layers L2-L6. Gliogenesis starts after E17. SVZ – subventricular zone, IP – intermediate progenitor, CR – Cajal-Retzius cell, BV – blood vessel, Ast – astrocyte, SEZ – subependymal zone, EL – Ependymal cell layer, DL Pyr – deep-layer pyramidal neuron, UL Pyr – upper layer pyramidal neuron, WM – white matter. Modified from (Kwan et al., 2012). (B) – Detailed schematic of neurogenesis in the cortex. For detailed description see text. Image taken from (Paridaen and Huttner, 2014).

All of those processes are regulated by an intricate network of highly conserved signaling pathways, the cell cycle, transcription factors, epigenetic modifications, and posttranscriptional mechanisms. Although a plethora of proteins (Taverna et al., 2014) and long noncoding RNAs (Aprea and Calegari, 2015) have been implicated in the regulation of corticogenesis no circular RNA was identified to participate in this complex process, so far.

As previously mentioned, most of our knowledge on neocortical neurogenesis stems from mouse. Mice however are a lissencephalic species, meaning that they lack gyri, or folds, on the cerebral cortex. Humans and primates on the other hand are gyrencephalic species, which results in a tremendous increase of the cortex' surface area which is often implied in higher cognitive functionality (Gautam et al., 2015). The reason for this gyrification could be the presence of two new cortical progenitor types whose pool is particularly increased in gyrencephalic species (Figure 6 - A). These so called outer subventricular zone (OSVZ) progenitors, include basal radial glial cells (bRGs) and basal radial glia-derived transit amplifying progenitors (Fietz et al., 2010; Hansen et al., 2010). Hence, although the insights gained from mice are invaluable, they cannot fully explain the more complex cortical development observed in humans and other primates.

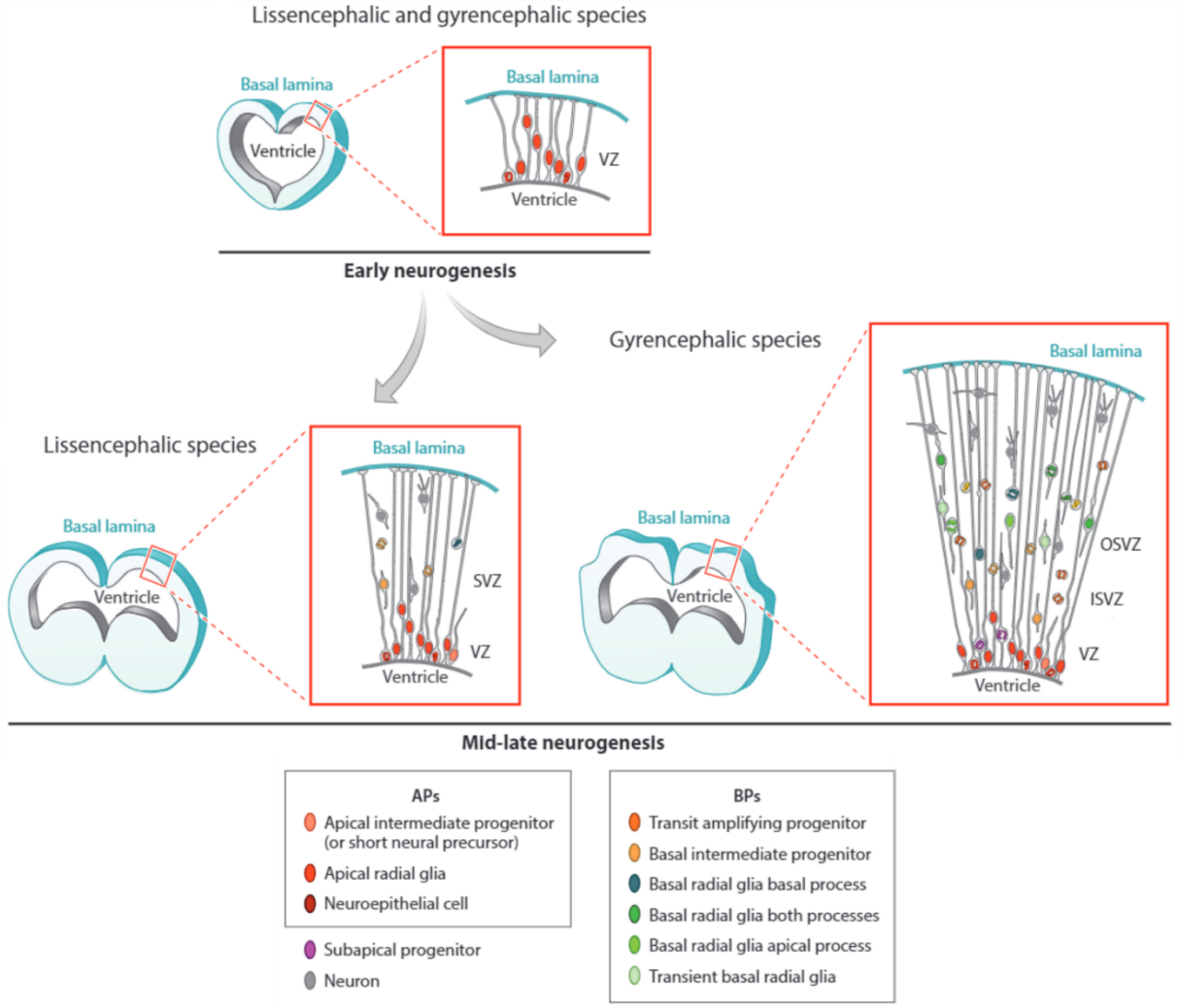


Figure 6: Differences in neurogenesis in mammalian neocortices depending on gyrification status.
Progenitor type diversity increased during evolution. Neuroepithelial cells are found in lissencephalic and gyrencephalic species. The variety and abundance of progenitor cells increases in mid and late neurogenesis in gyrencephalic species (e.g. humans, primates, ferret) compared to lissencephalic species (most rodents, lower vertebrates). (Taverna et al., 2014)

3.3 Investigating single-cell gene expression with spatial resolution

A cell is the basic biological unit of all living organisms. Often, researchers try to group them into cell types by their function, morphology, location, or by other criteria such as marker genes or molecular content. Grouping by molecular content, e.g. on RNA or protein level, is complicated by the measurement of the low amounts of RNA (10 - 30 pg/cell, (BioNumbers)) or protein per cell (42-675 pg/cell, (BioNumbers)). Hence, for a long time it was only possible to measure so called bulk RNA-seq samples which average over many, often different, cells (Figure 7 - A). Regardless of the importance of sequencing and analysis of bulk samples, important details might be lost due to the averaging over multiple cells and cell types. With the development of single-cell (q)PCR it became possible to investigate gene expression on the single-cell level for the first time, despite its very limited throughput (Bengtsson et al., 2008; Eberwine et al., 1992; Huang et al., 2014; Taniguchi et al., 2009; Warren et al., 2006). Single-cell RNA-seq followed soon (Tang et al., 2009), but collection of single cells by micropipetting or laser capture microdissection was tedious (Figure 7 - B). Furthermore, since each cell had to be processed separately those early approaches of single-cell sequencing were cost intensive.

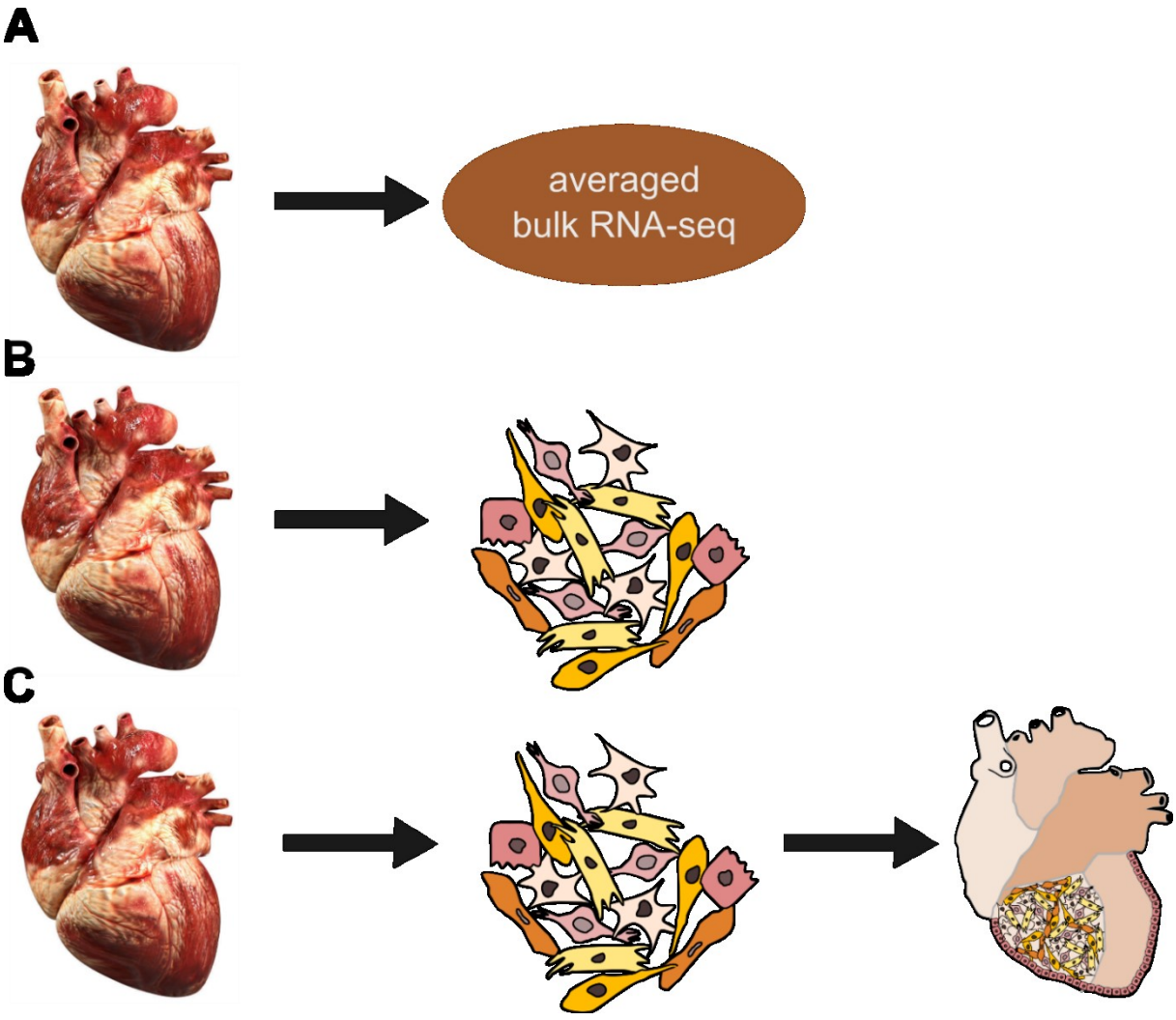
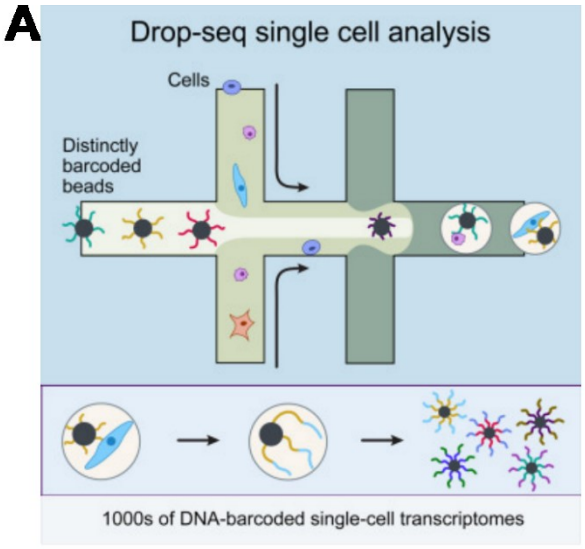


Figure 7: Advancements in resolution of RNA-seq from bulk to single-cell & spatial single-cell RNA-sequencing.
(A) – Bulk RNA-seq would average over all different cells and compartments of a human heart (human heart from cgtrader.com). **(B)** – Single-cell RNA-seq would allow the identification and investigation of single-cells from the same tissue. **(C)** – Recording spatial information before disintegration of the tissue would allow the investigation of single cells in their spatial context/neighborhood.

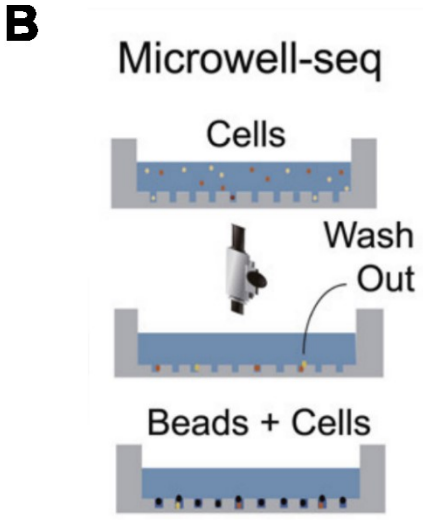
The development of single-cell preparation by microfluidics (e.g. Fluidigm C1) and microdroplets in conjunction with efficient enzyme-based tissue dissociation methods increased the throughput and lowered the costs of single-cell RNA sequencing tremendously. Since the onset of these single-cell sequencing methods a large number of papers has been published investigating cell populations and tissues at single-cell transcriptome resolution, providing insights into the cellular composition of highly complex tissues e.g. glioblastoma (Patel et al., 2014), mouse retina (Macosko et al., 2015) or the mouse cortex and hippocampus (Zeisel et al., 2015). A multitude of methods has been developed to efficiently prepare single cells for analysis, relying on droplets to encapsulate cells (Figure 8 - A) (Klein et al., 2015; Macosko et al., 2015), sorting single cells in individual wells (Figure 8 - B) (Han et al., 2018; Jaitin et al., 2014; Muraro et al., 2016) or labeling cells with barcode combinations to obtain single-cell resolution (combinatorial indexing) (Figure 8 - C) (Cao et al., 2017; Rosenberg et al., 2018).

However, all of these methods require that the tissue of interest is dissociated in the initial step. Thereby all spatial information is lost. To overcome this loss of information, gene expression can be directly probed on the tissue itself by smFISH and advanced variations of it to increase the number of detectable transcripts, e.g. by sequential hybridization rounds or fluorophore combinations
(Figure 8 - D) (Chen et al., 2015; Levesque et al., 2013; Lubeck et al., 2014) or direct *in situ* sequencing (Ke et al., 2013; Lee et al., 2014). Despite their high sensitivity, these methods are only able to detect a few hundred transcripts per cell that also have to be predefined in the used probe set. Furthermore, transcripts often cannot be assigned unambiguously to their cell of origin. A *co*-staining for a cell's outer membranes is not compatible with the protocol and in highly complex tissues like a mammalian cortex even staining of a cell's outer membrane would not allow to distinguish the tightly packed cells. Hence, only perinuclear or nuclear transcripts can be assigned to their cell of origin, limiting the quantitative power of highly multiplexed single molecular FISH.

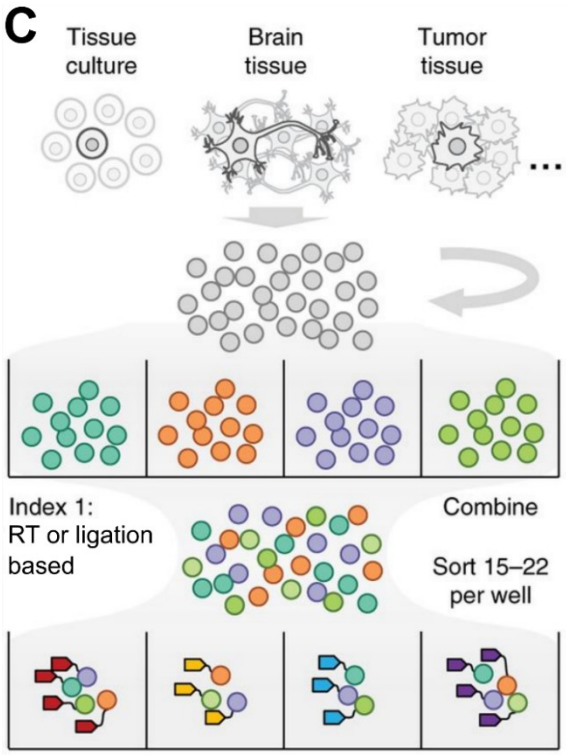
Therefore, tools were developed to complement single-cell sequencing data with spatial information. Using *in situ* experimental data to supplement single-cell RNA-sequencing data enabled the computational reconstruction at single-cell resolution, e.g. zebrafish embryo (Satija et al., 2015), the marine annelid *Platynereis dumerilii* (Achim et al., 2015) and the *Drosophila* embryo (Karaiskos et al., 2017). However, these methods relied on a pre-existing resource already holding highly reproducible spatial information for the tissue in question, a criterion that is often fulfilled in samples from model organisms but not in medically relevant, unique patient samples.



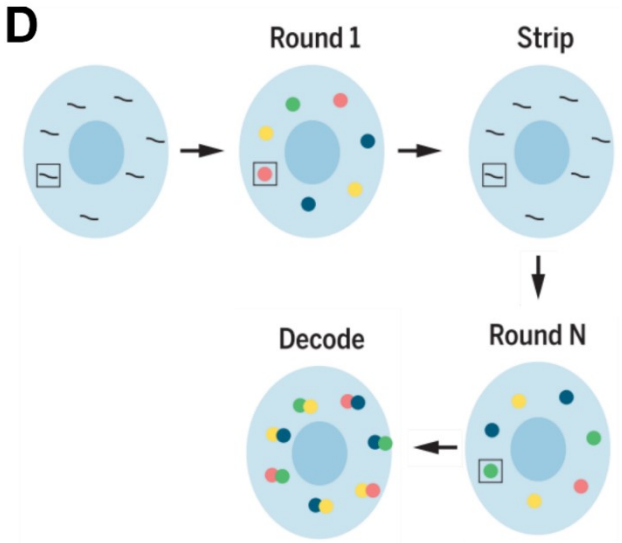
taken from Macosko et al., (2015)



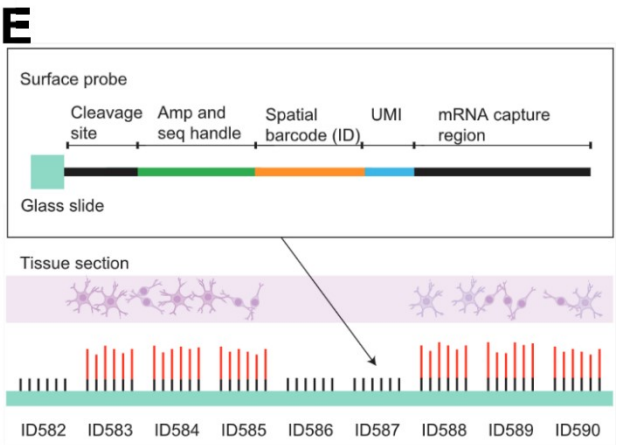
taken from Han et al., (2018)



taken from Vitak et al., (2017)



taken from Lein et al., (2017)



taken from Stahl et al., (2016)

Figure 8: Overview of various single-cell RNA-seq and RNA-seq methods with spatial resolution.
(A) – Drop-seq encapsulates single cells in nanodroplets together with a uniquely barcoded polyA-binding bead (Macosko et al., 2015). **(B)** – Microwell-seq traps single cells in nanowells before uniquely barcoded beads are added (Han et al., 2018). **(C)** – Combinatorial indexing relies on several rounds of barcoding. In-between these rounds cells are mixed and split. Each cell will hold a unique combination of barcodes in the end. (Cao et al., 2017; Rosenberg et al., 2018; Vitak et al., 2017). **(D)** – Principle of high-throughput single molecule FISH to determine spatial gene expression *in situ* (Lein et al., 2017). **(E)** - "Spatial transcriptomics" enables spatial gene expression measurements from tissue, averaging over 10-20 cells (Ståhl et al., 2016).

‘Spatial Transcriptomics’ (Ståhl et al., 2016) does not require additional information to produce a spatially resolved gene expression map. Here, the tissue is applied to a barcoded array and, depending on the cell density in the investigated tissue, the signal is averaged over 10-20 cells (Figure 8 - E). Again, using this setup single-cell resolution is lost. The combination of both, single-cell RNA sequencing and ‘Spatial Transcriptomics’, has been used to collect a tumor cell atlas from different tumor sections of a pancreatic ductal adenocarcinoma (Moncada et al., 2018). The highly complex cellular composition of a tumor, or many diseased tissues, makes an unambiguous mapping of cells onto a separately obtained gene expression map difficult, especially when relying on distinct tissue sections even when they originate from the same specimen. The following table summarizes current approaches (Table 2).

Table 2: Comparison of existing single-cell RNA-seq and spatial methods.

Method	Single-cell Resolution	Spatial Resolution	Limitations	Citation
Drop-seq	✓	**X**	No spatial resolution	(Macosko et al., 2015)
Tomo-seq	**X**	✓	Several highly reproducible specimens necessary	(Junker et al., 2014)
Spatial Transcriptomics	**X**	✓	No single-cell resolution	(Ståhl et al., 2016)
FISH-based methods	**✓ / X**	✓	Only (peri-)nuclear transcripts are assignable, limited target space, high turnaround time	(Chen et al., 2015; Femino et al., 1998; Lubeck and Cai, 2012; Lubeck et al., 2014; Raj et al., 2008)

Considering the advantages and drawbacks of the existing methods, there is a need for a method that allows to record spatial information of a single cell while still able to measure the transcriptome in an unbiased transcriptome-wide way. This method development is addressed in “7.3 - 3D-seq: a method to integrate single-cell and spatial transcriptomics”. The 3D-seq protocol will be the first approach that enables researchers to analyze the transcriptome of single cells within a complex tissue without losing the spatial information on the cell’s localization.

4 Aims of this thesis

1. **Collection of a circRNA expression catalogue in neural samples.**
    a. Elucidation of spatiotemporal expression patterns in neural tissues and during neuronal differentiation.
    b. Comparison of circRNA expression to corresponding mRNA expression.
    c. Analysis of circRNA conservation in human and mouse.
    d. Evaluation of ADAR1's influence on circRNA biogenesis in neural samples.

2. **Studying the function of circSLC45A4 in neuronal differentiation.**
    a. Analysis of circSLC45A4's expression pattern and conservation.
    b. Determining the localization, abundance, and expression changes in neuronal differentiation of circSLC45A4.
    c. Identification and analysis of the phenotype caused by circSLC45A4 perturbation (RNAi based) in human *in vitro* and mouse *in vivo* model systems.
    d. Identification of potential interaction partners – RNA-binding proteins or RNAs – by establishing a pulldown approach for endogenous circSLC45A4.
    e. Investigating the potential translation of circSLC45A4.

3. **Development of a method that allows single-cell RNA-seq while preserving information on the spatial localization of single cells within a tissue.**
    a. Establishing *in situ* reverse transcription in methanol-fixed tissue slices of adult mouse brain.
    b. Evaluating tissue digestion efficiency after the tissue was fixed with methanol and rehydrated.
    c. Optimization of a combinatorial indexing protocol on methanol-fixed tissue and in single cells.
    d. Design and production of a grid structure with CAD software and a 3D printer.
    e. Cell type identification and spatial reconstruction of an adult mouse brain slice *in silico.*

5 Material

5.1 Chemicals, enzymes and reagents

Chemical or reagent	Composition	Order #	Company
0.05 % Trypsin-EDTA	--------	25300062	Invitrogen
0.25 % Trypsin EDTA	--------	25200072	Invitrogen
10x CutSmart Buffer	50 mM Potassium Acetate 20 mM Tris-acetate 10 mM Magnesium Acetate 100 µg/ml BSA pH 7.9 @ 25°C	B7204S	NEB
10x MOPS Running Buffer	0.2 M MOPS, pH 7.0 (with NaOH) 50 mM Sodium Acetate 10 mM EDTA	M1254 S2889 E6758	Sigma Sigma Sigma
10x T4 DNA Ligase Buffer	50 mM Tris-HCl 10 mM $MgCl_2$ 1 mM ATP 10 mM DTT pH 7.5 @ 25°C	B0202S	NEB
32 % Para-formaldehyde	--------	15714-S	Science Service GmbH
Acrylamid-Bisacrylamid (37.5:1)	--------	3029.1	Roth
Agarose NEEO Ultra	--------	2267.3	Roth
Ammonium Persulfate	--------	A3678	Sigma
Benzonase	250 U/µl	71205-3	Millipore
Blocking Milk Solution	3 % (w/v) skim milk powder in PBS	70166	Sigma
Blocking Solution for Immunofluorescence	1 % (w/v) BSA 5 % (v/v) normal goat serum In PBS, pH 7.4	01400.100 ab7481	Biomol abcam
Blunt/TA Cloning Mix	--------	M0367S	NEB
BSA			
CDP-Star, ready-to-use	--------	12041677001	Roche
Collagenase, type IV	--------	17101015	Life Technologies
DIG RNA Labeling Mix, 10x	10 mM ATP, CTP, GTP 6.5 mM UTP 3.5 mM DIG-11-UTP	11277073910	Roche
dNTPs	100 mM dATP/dGTP/dCTP/dTTP	R0182	Fermentas
Dropseq Lysis Buffer	6 % (v/v) Ficoll PM-400 0.2 % (w/v) Sarkosyl 20 mM EDTA 30 mM Tris-HCl, pH 7.5 50 mM DTT	GE17030010 Roth
Dynabeads Streptavidin MyOne C1 beads	--------	65002	Life Technologies
EBSS	--------		
ECL Western Blotting Reagents	--------	RPN2109	GE Healthcare
Ethanol	--------	5054.1	Roth
EvaGreen 20x	--------	31000	Linaris
FastAP	--------	EF0651	Thermo Scientific
Ficoll PM-400	--------	GE17030010	GE Healthcare
Formaldehyde 37 %	--------	4779.1	Roth
GlycoBlue coprecipitant	15 mg/ml	AM9516	Life Technologies
Glyoxal Loading Dye (2x)	--------	AM8551	Ambion
Igepal CA-630	--------		

Isopropanol	--------	AE73.1	Roth
Lipofectamine 2000	--------	11668019	Life Technologies
Lipofectamine RNAiMax	--------	13778030	Life Technologies
Maxima H Minus reverse transcriptase	0.2 U/ml	EP0753	ThermoFisher
Maxima SYBR Green/ROX qPCR Master Mix	--------	K0223	Life Technologies
Methanol	--------	8388.2	Roth
Methanol	--------	50713	Lager
NorthernMax Pre-/Hybridization Buffer	--------	AM8677	Ambion
Papain	--------	LS003119	Worthington Biochemical Corp.
Paraformaldehyde for Fixation	4 % (w/v) paraformaldehyde In PBS, pH 7.4 (adjusted with HCl)	15714-S	VWR
PBS	137 mM NaCl 2.7 mM KCl 10 mM Na_2HPO_4 1.8 mM KH_2PO_4 pH 7.4, adjusted with HCl	3957.2 6781.1 T877.1 60347	Roth Roth Roth Sigma
PBST	1 % (v/v) Tween-20 in PBS	9127.2	Roth
PEG 8000	--------	1584	Roth
Pierce BCA Protein Assay Kit	--------	23225	ThermoFisher
Poly-L-lysine	1 % (w/v) poly-L-lysine In ddH_2O	P2636	Sigma
Proteinase K, recombinant, PCR grade, 20 mg/ml	--------	03115879001	Sigma Aldrich
Quick Ligation Kit	--------	M2200S	NEB
Random Hexamers	0.2 µg/µl	SO142	Fermentas
Ribolock	40 U/µl	EO0384	Fisher Scientific
RiboRuler high range RNA ladder	--------	SM1823	Fisher Scientific
RIPA Buffer	150 mM NaCl 1.0 % (v/v) NP-40 0.5 % (w/v) sodium deoxycholate 0.1 % (w/v) SDS 50 mM Tris-HCl, pH 8.0 Fresh: Complete mini protease inhibitor (EDTA-free)	3957.2 I3021 D6750 1057.1 4855.3 A32953	Roth Sigma Sigma Roth Roth ThermoFisher
SDS Loading Dye 5x (Laemmli)	10 % (w/v) SDS 25 % (v/v) beta-mercaptoethanol 50 % (v/v) glycerol 0.01 % (w/v) bromophenol blue 0.3125 M Tris-HCl, pH 6.8	1057.1 8057400005 G9012 B5525 4855.3	Roth Merck Millipore Sigma Sigma Roth
SDS-PAGE Running Buffer	25 mM Tris 190 mM Glycine 0.1 % (w/v) SDS	4855.3 39082 1057.1	Roth Roth Roth
SDS-PAGE Separation Buffer	1.5 M Tris-HCl, pH 8.8 0.4 % (w/v) SDS	4855.3 1057.1	Roth Roth
SDS-PAGE Stacking Buffer	0.5 M Tris-HCl, pH 6.8 0.4 % (w/v) SDS	4855.3 1057.1	Roth Roth
SSC Buffer 20x	--------	S6639	Sigma
Superscript II	--------	18064071	Invitrogen
Superscript III	--------	18080085	Invitrogen
Superscript IV	--------	18090200	Invitrogen

SYBR Gold Nucleic Acid Gel Stain	--------	S11494	MoBiTec GmbH
T4 DNA Ligase	--------	M0202L	NEB
T4 PNK	--------	EK0032	Fisher Scientific
T7 RNA Polymerase	--------	10881775001	Roche
TEMED	--------	2367.3	Roth
Terra PCR Direct Polymerase mix	--------	639271	Clontech
Transfer Buffer	25 mM Tris-HCl, pH 8.5 190 mM Glycine 0.05 % (w/v) SDS 20 % (v/v) Methanol	4855.3 39082 1057.1 8388.2	Roth Roth Roth Roth
Triton X-100	--------		
Triton X-100 (various concentrations v/v)	--------	T8787-50ML	Sigma
TRIzol	38 % (v/v) Phenol, pH 4.5 800 mM Guanidiniumthiocyanat 400 mM Ammoniumthiocyanat 100 mM Sodium acetate, pH 5.0 (adjust pH with acetic acid) 5 % (v/v) Glycerin	A980.1 2628.4 4477.4 6773.2 3783.2	Roth Roth Roth Roth Roth
TrypLE	--------	12604013	Life Technologies
Vectashield Hardset Mounting Medium with DAPI	--------	H-1500	Linaris

5.2 Consumables

Chemical or reagent	Order #	Company
10 cm dish, cell culture coated	83.3902	Sarstedt
12-well plate, cell culture coated	83.3921	Sarstedt
15 cm dish, cell culture coated	83.3903	Sarstedt
24-well plate, cell culture coated	662160	Greiner
6-well plate, cell culture coated	83.3920	Sarstedt
96-well Multiply FAST PCR plate	72.1981.202	Sarstedt
96-well Multiply FAST PCR plate (qPCR)	721981202	Sarstedt
BZO Seal Film, Adhesive, Optical Film	712350X	Biozym
Coverslips		
DNA LoBind tubes, 1.5 ml	5250130	VWR
Falcon tubes, 15 ml	50199	Lager
Falcon tubes, 50 ml	50200	Lager
Filter pipet tips SafeSeal SurPhob, 10 µl	VT0200	Biozym
Filter pipet tips SafeSeal SurPhob, 1250 µl	VT0270	Biozym
Filter pipet tips SafeSeal SurPhob, 20 µl	VT0220	Biozym
Filter pipet tips SafeSeal SurPhob, 300 µl	VT0250	Biozym
Hybond N+ membrane	RPN303B	VWR
Hybond-P PVDF membrane	GE10600023	Sigma
MAXYMUM Recovery tubes	11311984	Fisher Scientific
Microamp 8-well caps	N8010535	Thermo Scientific
Microamp 8-well strips	N8010580	Thermo Scientific
PCR strips, 0.2 ml	72991002	Sarstedt
PLA/PHA filament mix	CF551053	Reichelt
Superfrost Plus Adhesion Microscope Slides	J1800AMNT	Thermo Scientific
Twintec LoBind PCR plates	0030129504	Neolab

5.3 Commercially available kits

Kit	Order #	Company

AMPure RNA clean beads	29168	Beckmann Coultier
AMPure XP beads	A63881	Beckmann Coultier
BaseScope Reagent Kit	322900	ACDBio
Bioanalyzer High sensitivity DNA Chip kit	50674626	Agilent
Bioanalyzer RNA Nano 6000 Chip Kit	50671511	Agilent
Bioanalyzer RNA Pico 6000 Chip Kit	5067-1513	Agilent
ERCC RNA spike-in mix	4456740	Life Technologies
Mix&Go Competent Cells, DH5α	T3009	Zymo Research
Nextseq 500 High Output Kit v2 (150 cycles)	FC-404-2002	Illumina
Nextseq 500/550 Mid Output Kit v2 (150 cycles)	FC-404-2001	Illumina
Qubit dsDNA HS Assay Kit	Q32854	Invitrogen
RiboMinus Eukaryote v2	A15026	Life Technologies
Tapestation High sensitivity D5000 Ladder	50675594	Agilent
Tapestation High sensitivity D5000 Reagents	50675593	Agilent
Tapestation High sensitivity D5000 Screen Tape	50675592	Agilent
Truseq RNA Sample Prep Kit v2	RS-122-2001	Illumina
Truseq Small RNA Sample Prep Kit A	RS-200-0012	Illumina
Truseq Stranded mRNA Library Prep kit	RS-122-2102	Illumina
Zymoclean DNA Gel Recovery	D4007	Zymo Research
ZymoPURE Plasmid Midiprep Kit	D4201	Zymo Research
ZymoPURE Plasmid Miniprep Kit	D4212	Zymo Research

5.4 Devices

- 96-well magnetic rack, DynaMag (Life Technologies, 12027)
- ABI StepOne Plus (ThermoFisher)
- Bioanalyzer 2100 (Agilent, G2939BA)
- Brand Transferpette S, 0.5-10 µl / 2-20 µl / 20-200 µl / 100-1000 µl
- Eppendorf Centrifuge 5424R
- eppendorf MixMate
- Eppendorf X50s Mastercycler (Eppendorf, EPPE6311000.010)
- EVE cell counter (VWR, 7342675)
- GeneTouch Thermal Cycler (Biozym, 685012)
- Keyence BZ-X700 microscope with 4x (CFI Plan Apo λ, 972030), 10x (CFI Plan Apo λ, 972031), 20x (CFI Plan Apo λ, 972032), 40x (CFI Plan Apo λ, 972033), 60xOil (CFI Plan Apo λ, 972036) objectives
- LAS4000 (Fujifilm)
- Mini-PROTEAN Tetra Cell System (BIO-RAD)
- Nanodrop 1000 (Thermo Scientific)
- neoLab IntelliMixer
- Nikon TiE Inverted microscope with 4x (Plan Fluor, MRH20041), 10x (Plan Apo λ, MRD00105), 20x (Plan Apo λ, MRD00205), 20xELWD (S Plan Fluor, MRH48230), 40xOil (CFI Plan Fluor, MRH01401), 60x Oil (Plan Apo λ, MRD01605) objectives and Andor iXonUltra888 EMCCD camera
- OMEGA Pipettor Plus
- Owl EasyCast Gel Electrophoresis System (ThermoFisher)
- Peqstar 2x Thermocycler (Peqlab/VWR, 732-2888)
- Projet HD 3000
- Qubit 4 (Thermo Scientific, Q33226)
- Roth table top centrifuge

- TapeStation 4200 (Agilent, G2991AA)
- Trans-Blot SD Semi-Dry Electrophoretic Transfer Cell (BIO-RAD)
- ultimaker 2+
- Vortex Genie 2
- Zeiss AxioObserver Z1 with AxioCam MRm (Zeiss)

5.5 Plasmids

Name	Purpose	Size
pSilencer2.1-shRNA_circSlc45a4_1	circSlc45a4 knockdown *in utero*	5.1 kb
pSilencer2.1-shRNA_circSlc45a4_2	circSlc45a4 knockdown *in utero*	5.1 kb
pSilencer2.1-shRNA_circSlc45a4_3	circSlc45a4 knockdown *in utero*	5.1 kb
pSilencer2.1-shRNA_mRNA_Slc45a4_1	mRNA Slc45a4 knockdown *in utero*	5.1 kb
pSilencer2.1-shRNA_mRNA_Slc45a4_2	mRNA Slc45a4 knockdown *in utero*	5.1 kb
pSilencer2.1-shRNA_scramble	Control knockdown *in utero*	5.1 kb

5.6 Antibodies

Antibody	Dilution	Order #	Company
Anti-ADAR1	1:1000	ab168809	abcam
Anti-Digoxigenin-AP, Fab fragments	1:20000	11093274910	Roche
Anti-DOCK7	1:1000	ab118790	abcam
Anti-FLAG	1:2000	F3165-1mg	Sigma
Anti-FXR2	1:1000	A303-893A	Biomol
Anti-GAPDH	1:5000	G8795-200UL	Sigma
Anti-GFAP	1:1000	AB5804	Merck/Millipore
Anti-GFP	1:400	600-1-1-215	Rockland
Anti-Reelin	1:100	MAB5364	Merck/Millipore
Anti-SARM1	1:1000	13022S	Cell Signaling Technologies
Anti-Tbr2	1:500	ab183991	abcam
Anti-TUBB3	1:1000	801213	BioLegend
Goat Anti-Mouse 488	1:2000	ab150113	abcam
Goat Anti-Rabbit 647	1:2000	ab150079	abcam
Polyclonal Goat anti-Mouse HRP	1:5000	P0447	Dako
Polyclonal Goat anti-Rabbit HRP	1:5000	P0448	Dako

6 Methods

6.1 Biochemical and Molecular Biology methods

6.1.1 Animals

For results chapter 7.2 - CircSLC45A4 is required to keep neuronal cells in a progenitor state in cell culture systems and in the mammalian brain:

Plugged C57BL/6J females were purchased from Janvier Labs and housed in the Biomedical Services facility of the MPI-CBG under standard conditions. All experimental procedures were approved by the “Landesdirektion Sachsen” (license TVV 16-2018).

For results chapter 7.3 - 3D-seq: a method to integrate single-cell and spatial transcriptomics:

C57BL/6J animals were housed in the MDC animal facility under standard conditions. All experimental procedures were approved by “LaGeSo Berlin” (license T0325/10).

6.1.2 BaseScope

The BaseScope assay by ACD can be used to detect single splice junctions. The signal strength is increased by using a branched DNA detection system (ACDBio, 2017). In brief, embryonic mouse brain was fixed with 4 % PFA overnight, cryoprotected overnight with 30 % sucrose, and then frozen in tissue freezing medium. Then slides were cryosectioned at 10 µm and carried through the BaseScope standard protocol for fixed frozen tissue sections, except that proteinase digest was limited to 10 min at room temperature to avoid over digestion.

Slices were mounted with one drop of Vectashield Hardset Mounting Medium with DAPI and imaged using a Keyence BZ 9000 microscope.

6.1.3 Cell fractionation

For cell fractionation a kit provided by abcam was used (ab109719). RNA was extracted from cytoplasmic, nuclear and mitochondrial fractions with TRIzol and analyzed by RT-qPCR. Sufficient fractionation of the cells was monitored by using ACTB and GAPDH as positive control for the cytoplasmic fraction, XIST and NEAT1 for the nuclear fraction and MT-CO2 for the mitochondrial fraction.

6.1.4 Cloning, transformation and propagation of plasmids

Cloning was performed according to standard methods. Plasmids were cut using appropriate restriction enzymes, dephosphorylated with FastAP and purified from an agarose gel with a Zymoclean Gel DNA Recovery kit. Inserts were either PCR amplified and then digested with the same set of restriction enzymes or for cloning of shRNAs, oligonucleotides with fitting overhangs were annealed and phosphorylated with T4 PNK. Insert and cut plasmid were ligated with T4 DNA Ligase and subsequently transformed into chemically competent bacteria (DH5α or Mix&Go Competent Cells, DH5α).

The ZymoPURE Plasmid Miniprep kit, or the ZymoPURE Plasmid Midiprep kit were used for minipreps or midipreps of bacterial DNA, respectively.

6.1.5 Immunofluorescence

Immunofluorescence staining was performed according to standard protocols. In brief, cells were grown on poly-L-lysine (PLL) coated coverslips. Adherent cells were washed three times with PBS, fixed for 20 min with 4 % (w/v) paraformaldehyde (PFA) at room temperature and then washed again three times with cold PBS. Cells were permeabilized with PBST (15 min, 4°), again washed twice with PBS and then blocked with blocking solution (1 h, 4°C). Then the primary antibody was applied, diluted in PBS with 1 % (w/v) BSA, and incubated at 4°C overnight. Unbound antibody was washed away by applying cold PBS 4 times. The secondary antibody was diluted in blocking solution and applied for 2 h at 4°C in the dark to avoid fluorophore bleaching. Again, unbound antibody was washed away with three PBS washing steps. To mount the coverslips on microscope slides for imaging, 10 µl Vectashield Mounting Medium with DAPI were used per coverslip. Mounting medium was dried at room temperature for up to 1 h and then coverslips were fixed on the microscope slide with clear nail polish.

Slides were imaged using either a Zeiss AxioObserver Z1 with an AxioCam MRm attached or a Keyence Bx-700.

6.1.6 Immunofluorescence staining of cortical slices & quantification

For immunohistochemistry, brains were fixed overnight at 4˚C in 4% paraformaldehyde (in 0.1M phosphate buffer), cryoprotected in 30% sucrose and cryosectioned (10 µm thick). For antigen-retrieval cryosections were incubated for 1 h at 70˚C (in 0.01 M Citrate Buffer, pH=6.0), followed by permeabilization for 20 min (in 0.5 % Triton X-100), quenching for 30 min (in 0.1 M glycine) and blocking for 1h (10 % donkey serum 0.3 % Triton X-100) at RT. All primary antibodies were incubated overnight at 4˚C (in 3 % donkey serum, 0.3 % Triton X-100), whereas secondary antibodies for 2 h at room temperature. Nuclei were counterstained with DAPI. Imaging was performed using an ApoTome fluorescence microscope (Carl Zeiss) and pictures were assembled using Axiovision software (Carl Zeiss). For quantifications, composites were manually analyzed using Photoshop CS6 (Adobe) and sample averages compared by One-way ANOVA. p-values < 0.05 were considered significant.

GFP antibody – Rockland 600-1-1-215; 1:400

Secondary antibodies – Jackson Laboratory; 1:500

6.1.7 *In utero* electroporation & FACS-sorting of GFP+ cells

Plasmids for *in* utero electroporation (IUE) were generated by replacement of the Hygromycin cassette of pSilencer2.1-U6-hygro (Thermo Fisher) with a nuclear-localized GFP, followed by insertion of shRNAs (purchased as synthetic oligonucleotides from Eurofins). IUE was performed as previously described (Artegiani et al., 2012). Briefly, C57BL/6J E13.5 pregnant mice were anesthetized with Isoflurane (Baxter) and 2-3 µg of plasmid DNA were injected into the embryo ventricle, followed by the application of 6 electric pulses (30V and 50 ms each at 1 s intervals) through platinum electrodes using a BTX-830 electroporator (Genetronics). Embryo brains were collected at E 15.5 for immunochemistry (whole brain) or FACS-sorting of electroporated cells (GFP+) (lateral cortices). After removal of meninges and ganglionic eminences, cortices were dissociated with papain-based Neural Tissue Dissociation kit

(Miltenyi Biotech) and cells were resuspended in 500 µl of ice-cold PBS. DAPI (1:1000) was added for dead cells discrimination and sorting of GFP+ cells was performed by BD FACSAriaTM III (BD Biosciences).

6.1.8 Isolation of synaptoneurosomes from mouse brain

Synaptoneurosomes, a fraction enriched in pinched-offed resealed presynaptic terminals still attached to resealed postsynaptic processes, were isolated as previously described (Gray and Whittaker, 1962; Huttner et al., 1983). Briefly, one adult NMRI mouse brain was dissected and immediately homogenized in 10 volumes of ice-cold homogenization buffer (0.32 M sucrose in 10 mM Hepes, pH 7.4) supplemented with Complete Protease Inhibitor cocktail (Roche) and 160 U/ml SUPERase-In RNase inhibitor (Ambion). All the following steps were performed at 4°C: The homogenate was centrifuged (1000 x g, 10 min and 1500 x g, 10 min) to remove nuclei and cellular debris. The supernatant was further centrifuged at 17,500 x g for 30 min and the pellet (P2) was dissolved in homogenization buffer. P2 was layered on top of a 0.8/1.2 M sucrose density gradient and centrifuged at 230,000 x g in a swinging bucket rotor for 20 min. The synaptoneurosomes fraction was recovered from the 0.8/1.2 M interphase band, layered on top of 0.8 M sucrose and centrifuged as above. The pellet (P2') was dissolved in homogenization buffer and used for further analysis.

6.1.9 Northern Blotting and probe generation

TRIzol-extracted total RNA (10-50 µg) was denatured by adding Glyoxal Loading Dye and incubating the samples for 30 min at 50°C. Denatured RNA was loaded on a denaturing agarose gel (1.2 % (w/v) agarose in MOPS buffer, 3 % (v/v) formaldehyde) and run at 80 V until intended separation is achieved. The gel was stained with SYBR Gold and imaged with a Fujifilm LAS4000 gel imaging system. The agarose gel was placed onto a wetted Amersham Hybond N+ nitrocellulose membrane and put between two with 1x TBE soaked Whatman papers. Nucleic acids were transferred onto the membrane was achieved by applying an electrical current in a semi-dry blotting apparatus for 1 hour at 15 V. The membrane was dried, and proper transfer was checked by imaging the membrane in the EtBr channel of an LAS4000. The membrane was washed with 50 mM Tris, pH 8.0 for 10 min at 60°C, prehybridized in NorthernMax Pre-/Hybridization Solution for 30 min at 68°C and hybridized with a specific DIG-labeled probe (0.1 nM, denatured: 2 min at 98°C) overnight at 68°C. Unbound or unspecifically-bound probe was washed away with 2x SSC, 0.1 % (w/v) SDS, the solution was changed after 30 min at 68°C for 3 times. Two additional washing steps with 0.2x SSC, 0.1 % SDS for 30 and 60 min at 68°C further increased the specific signal. Detection was achieved by following the manufacturer's protocol of CDP Star ready-to-use detection solution and imaging in a Fujifilm LAS4000.

Probes used for Northern Blotting were approximately 150 nt in length. The region to be targeted by the probe was amplified by PCR at which point the sequence of the T7 promoter was inserted (5'-TAATACGACTCACTATAGGGAG-3'). The T7 promoter was then employed during *in vitro* transcription (IVT) to amplify an RNA probe for northern blot detection. Detectability of the probe was achieved by incorporating digoxigenin-11-UTP into the RNA probe, by using DIG RNA Labeling Mix according to the manufacturer's instruction.

6.1.10 Preparation of polyA+ libraries and sequencing

PolyA+ libraries were prepared from RNA with Illumina TruSeq RNA Library Prep kit v2 (SH-SY5Y) or Illumina TruSeq Stranded mRNA LT (mouse cortex GFP+ cells) and sequenced on an Illumina NextSeq 500 with 1x150 bp.

6.1.11 Preparation of small RNA libraries and sequencing

Small RNA libraries were prepared from total RNA with an Illumina TruSeq Small RNA Sample Prep Kit and sequenced on a NextSeq 500 with 1x50 bp.

6.1.12 Preparation of total RNA libraries and sequencing

2 µg of total RNA were used as starting material for total RNA libraries and spiked with ERCC RNA spike-in mix. Then, ribosomal RNA (rRNA) was depleted with the Ribominus Eukaryote Kit v2 or with a RNase H approach (Adiconis et al., 2013). Successful ribodepletion was assessed using a Bioanalyzer RNA 6000 Pico Chip. RNA was fragmented and subsequently prepared for sequencing using an Illumina Truseq Unstranded Total RNA Library Prep kit or an Illumina Truseq Stranded Total RNA Library Prep kit. Libraries were sequenced on a Hiseq 2000 or Nextseq 500, at 1x150 nt.

6.1.13 Protein Extraction and Western Blotting

In general, adherent cells were harvested by scraping and centrifugation (500 x g, 5 min, 4°C), washed once with PBS and resuspended with twice the volume of the cell pellet in ice-cold lysis buffer (RIPA buffer, see 5.1). Cells were incubated on ice for 30 min, while the lysate was passed through a 21G needle 6 times. The lysate was cleared by centrifugation (16000 x g, 20 min, 4°C) and the supernatant was kept for SDS-Polyacrylamide gel electrophoresis and western blot analysis.

10 to 20 µg of protein, as determined by BCA assay, were denatured by addition of 5x SDS Loading dye and incubation at 95°C for 5 min. Proteins were then separated by size by SDS-PAGE (Laemmli, 1970). Subsequent transfer of proteins onto a methanol activated PVDF membrane was achieved by semi-dry blotting in Transfer buffer. The membrane was blocked for 30 min at room temperature with 3 % (w/v) skim milk in PBS and incubated with protein specific primary antibody overnight at 4°C, slowly shaking. Unbound primary antibody was washed: 3 times 10 min in PBST, slowly shaking. Secondary antibody was incubated for 2 h at room temperature, slowly shaking, and again unbound secondary antibody was washed away. Detection was achieved by addition of an HRP substrate (Amersham ECL) and the signal was quantified with a Fujifilm LAS4000.

6.1.14 Pulldown of circular RNA and mass spectrometric analysis

This protocol is mainly based on RAP-MS (Engreitz et al., 2015) and (Theil et al., submitted 2018). First, cells were grown exponentially and then interacting RNAs and protein were crosslinked at 254 nm (600 mJ/cm^2) and harvested by scraping in 1xPBS. Cells were washed in 1xPBS, pelleted (500xg, 5 min) and stored at -20°. Next, 30x10^6 cells per sample were lysed by adding 900 µl Total Cell Lysis Buffer that was supplemented with Protease Inhibitor (1x complete Mini protease inhibitor) and RNase inhibitor (Ribolock 0.08 U/µl). Cells were lysed

on ice for 10 min and during that time passed 5 times through a 26G needle. Subsequently, the lysate was treated with RQ DNase I (2 U/ml) and RNase R (40 U/ml) for 15 min at 37°C in a thermomixer. The reaction was stopped by adding RQ DNase I Stop Solution (50 µl/ml). Then, the samples were prepared for pulldown by mixing them with the 2-fold volume of 1.5x Hybridization buffer and incubation on ice for 10 min. Debris was pelleted (16000xg, 4°C, 10 min) and the supernatant was incubated with 450 pM denatured probes specific for circSLC45A4 the first exon of mRNA SLC45A4 or negative control probes, specific to GFP, for 3 h at 37°C (thermomixer, shaking at 700 rpm, samples inverted every 30 min). Then Streptavidin MyOne C1 magnetic beads were prepared (150 µl per sample) by washing them 4 times with 10 mM Tris-HCl, pH 7.5, 2 times with 1x Hybridization Buffer and resuspension in 50 µl/per sample 10 mM Tris-HCl, pH 7.5. 50 µl prepared Streptavidin beads were added to each sample and incubate for 1 h at 37°C while mixing at 700 rpm. Then a magnetic rack was used to separate the beads from the flow-through and beads were washed 4 times with 500 µl of 1x Hybridization Buffer and 3 times with Washing Solution (Hybridization Buffer without detergents). The last washing step was used to split the sample 4:1 for protein and RNA detection. For protein detection, the beads were resuspended in 100 µl Protein Elution buffer with benzonase (100 U) and incubated for 3 h at 37°C, shaking at 1300 rpm. Supernatant was kept for protein identification by mass spectrometry. For RNA detection, beads were resuspended in 100 µl Proteinase K solution (1 mg/ml Proteinase K) and incubated at 50°C for 45 min, shaking at 1300 rpm, before 1 ml TRIzol was added for subsequent RNA extraction.

Sample preparation for proteome analysis

Protein samples were resuspended in 50 µL of 8 M urea and 0.1 M Tris-HCl, pH 8. Proteins were reduced with 10 mM dithiothreitol (DTT) at room temperature for 30 min and alkylated with 50 mM iodoacetamide (IAA) at room temperature for 30 min in the dark. Proteins were first digested by lysyl endopeptidase (LysC) (Wako) at a protein-to-LysC ratio of 100:1 (w/w) at room temperature for 3 hr. Then, the sample solution was diluted to a final concentration of 2 M urea with 50 mM ammonium bicarbonate (ABC). Trypsin (Promega) digestion was performed at a protein-to-trypsin ratio of 100:1 (w/w) under constant agitation at room temperature for 16 hr. Peptides were desalted with C18 Stage tips (Rappsilber et al., 2007) prior to LC-MS/MS analysis.

LC-MS/MS analysis

Reversed-phase liquid chromatography was performed by employing an EASY nLC 1200 (Thermo Fisher) using self-made fritless C18 microcolumns (Ishihama et al., 2002) (75 µm ID packed with ReproSil-Pur C18-AQ 3-µm resin, Dr. Maisch GmbH) and connected on-line to the electrospray ion source (Proxeon) of a Q Exactive plus (Thermo Fisher). The mobile phases consisted of (A) 0.1% formic acid and 5% acetonitrile and (B) 0.1% formic acid and 80% acetonitrile. Peptides were eluted from the analytical column at a flow rate of 200 nl/min by altering the gradient: 5-6% B in 2 min, 6-8% B in 14 min, 8-20% B in 44 min, 20-33% in 50 min, 33-45% B in 12 min, 45-60% B in 2 min, 60-95% B in 1 min. A Q Exactive (plus) instrument

was operated in the data dependent mode with a full scan in the Orbitrap followed by top 10 MS/MS scans using higher-energy collision dissociation (HCD). For standard proteome analyses, the full scans were performed with a resolution of 70,000, a target value of $3x10^6$ ions and a maximum injection time of 20 ms. The MS/MS scans were performed with a 17,500 resolution, a $1x10^6$ target value and a 20 ms maximum injection time. Isolation window was set to 2 and normalized collision energy was 26. Ions with an unassigned charge state and singly charged ions were rejected. Former target ions selected for MS/MS were dynamically excluded for 30 s.

Processing of mass spectrometry data

All raw data were analyzed and processed by MaxQuant (v1.5.1.2) (Cox and Mann, 2008). Default settings were kept except that 'match between runs' was turned on. Search parameters included two missed cleavage sites, cysteine carbamidomethyl fixed modification, and variable modifications including methionine oxidation, protein N-terminal acetylation, deamidation of glutamine and asparagine. The peptide mass tolerance was 6 ppm and the MS/MS tolerance was 20 ppm. Database search was performed with Andromeda (Cox and Mann 2008; Cox et al. 2011) against UniProt/Swiss-Prot human or mouse database (downloaded on 2014-11) with common serum contaminants and enzyme sequences. False discovery rate (FDR) was set to 1% at peptide spectrum match (PSM) level and at protein level. A minimum peptide count required for protein quantification was set to two. A label free quantification (LFQ) method implemented in MaxQuant (Cox et al., 2014) was used to quantify proteins. First, we only selected proteins that were quantified in at least 2 out of 3 biological replicates and missing LFQ values were replaced by imputation with a default setting in Perseus (Tyanova et al., 2016). Then, fold-changes and corrected p-values (two sample t-test) were computed for the volcano plot.

6.1.15 Quantification of neurite length

Fiji (Schindelin et al., 2012) was used to quantify neurite length on brightfield microscopy images taken at 20x with a Keyence BZ 9000. Therefore, the freehand line tool was used to trace neurites starting from the cell body to the neurite tip. The measure tool was used to determine the length in µm. 30 cells per condition were counted.

6.1.16 Quantitative real time PCR (qPCR)

All quantitative real time PCR experiments were conducted on an ABI StepOne Plus instrument. Depending on the sample and target gene set, cDNA was diluted 1:10 to 1:100 with ddH_2O and mixed with 2x Maxima SYBR Green/ROX qPCR Master Mix and 3.75 µl of 2.5 µM Primer mix to obtain 15 µl of total reaction volume per well. All measurements were conducted at least in technical triplicates.

For miRNA expression analysis, TaqMan assays with an attached FAM dye were quantified using TaqMan Universal master Mix II, no UNG-1.

Relative quantification was performed by using the comparative $\Delta\Delta C_T$ method (Pfaffl, 2001).

6.1.17 Reverse Transcription (RT) of RNA

If not stated differently, RNA was reverse transcribed to cDNA using Maxima H Minus Reverse Transcriptase at 100 U per µg RNA. First, up to 1 µg of RNA was mixed with 0.5 mM dNTPs, 0.25 µg random hexamers and filled up to 15 µl with double-distilled water (ddH_2O). This mixture was denatured at 65°C for 5 min in a thermocycler, immediately placed on ice and 100 U/µg RNA of Maxima RT, 4 µl 5x RT Buffer and 0.5 µl Ribolock were added. The cDNA synthesis was carried out with the following program in a thermocycler: 10 min at 25°C, 30 min at 50°C, 5 min at 85°C.

6.1.18 Reverse Transcription of miRNAs

The reverse transcription of miRNAs is more complex than standard total RNA reverse transcription due to their short size, typically around 22 nt, and the lack of a polyA-tail. Therefore, stem loop primers are used that only need a short overlap with the miRNAs 3'end. 50 ng of total RNA per target miRNA or housekeeper are reverse transcribed using SuperScript III (100 U/reaction), 0.5 mM dNTPs, 20 U Ribolock, 1x TaqMan RT primer and 1x first-strand synthesis buffer were filled up to 20 µl reaction volume with ddH_2O and mixed. Then the reaction was incubated for 30 min at 16°C, 30 min at 42°C and terminated for 5 min at 85°C.

6.1.19 RNA extraction

RNA was generally extracted using self-made TRIzol (see 5.1). Adherent cells were washed once with 1x PBS pH 7.4 and TRIzol was directly added. The samples were incubated for 5 min at room temperature to facilitate complete dissociation. Then 200 µl of Chloroform were added, samples were shaken vigorously and immediately centrifuged (15 min, 12 000xg, 4°C). The upper aqueous phase was recovered and contained RNA was precipitated by adding the same volume of isopropanol. Samples were inverted and incubated for at least 10 min at -20°C. To further aid precipitation of samples with low RNA content, 0.5 µl Glycogen (GlycoBlue) were added and samples were incubated for up to 16 h at -20°C before pelleting of the RNA and to further aid precipitation of RNA. Pelleting of the RNA was achieved by centrifugation (20 min, 16 000xg, 4°C). The supernatant was discarded, and the RNA pellet was washed twice with 80 % (v/v) ethanol. The pellet was dried (5 min, room temperature) and resuspended in water.

6.1.20 RNA interference

RNAi based knockdown of circSLC45A4 in SH-SY5Y cells was performed according to standard protocols with Lipofectamine RNAiMax (ThermoFisher). Briefly, $2x10^4$ cells/cm^2 were seeded in a 6-well plate the day before transfection. 25 pmol of siRNA were used to achieve knockdown of the intended target. SH-SY5Y cells were harvested 96 h post-transfection.

6.1.21 RNase R treatment

In order to degrade linear RNA, total RNA was treated with 3 U/µg RNase R for 15 min at 37°C in 1x RNase R buffer. The same amount of total RNA was used for a mock treatment without RNase R. After the reaction RNA from *C. elegans* was used as a spike-in at 10 % of the original RNA amount (for human samples). Then phenol/chloroform/isoamylalcohol was added to purify the remaining RNA. Therefore, the mixture was vigorously shaken on a Vortex Genie 2

and centrifuged (12000xg, 10 min, 4°C) to separate the phases. The aqueous phase containing the RNA was recovered and RNA was precipitated by addition of the threefold volume of ethanol, incubation on ice for 10 min and centrifugation (16000xg, 15 min, 4°C). The RNA pellet was washed with 80 % (v/v) ethanol and then the RNA was used for reverse transcription and analysis by qPCR. For normalization the previously added *C. elegans* spike-in RNA was used.

6.1.22 Stellaris single-molecule FISH

Stellaris single molecule FISH was performed according to the manufacturer's protocols (Biosearch Technologies). For circSLC45A4, 29 probes were used, and for mRNA SLC45A4 as well. As a negative control, samples were carried through the protocol without addition of probes.

The sequences of all Stellaris probes used can be found in 10 - Appendix: 10.1 - Oligonucleotides.

6.2 Experimental methods related to 3D-seq

6.2.1 *In situ* reverse transcription

In situ reverse transcription of cells was based on (Cao et al., 2017; Rosenberg et al., 2018). In brief, cells were mixed with reverse transcription mix:

Maxima H- RT		Superscript IV RT	
Reagent	Volume [µl]	Reagent	Volume [µl]
100 µM RT primer	1.0	100 µM RT primer	1.0
10 mM dNTPs	1.0	10 mM dNTPs	1.0
Cell suspension @ $1x10^5$ cells/ml	10.0	Cell suspension @ $1x10^5$ cells/ml	10.0
5x Maxima RT buffer	4.0	5x Superscript IV buffer	4.0
Ribolock	0.5	Ribolock	0.5
Ficoll PM-400 (20 %)	3.0	100 mM DTT	1.0
Maxima H- RT enzyme	0.5	water	2.0
		Superscript IV enzyme	0.5

Maxima H- RT and Superscript IV were incubated for 30 min at 25°C, 30 min at 37°C and 60 min at 42°C.

For optimizing conditions of Superscript IV RT, several additives were used as follows:

- Proteinase K: 0.08 µg/ml final concentration of Proteinase K are incubated with cells for 10 min at room temperature, shaking at 700 rpm. Reaction is stopped by adding 600 µl PBS+ 0.01 % (w/v) BSA. Cells are pelleted (3000xg, 10 min, 4°C) and washed with PBS.

- Ficoll PM-400: 5 % (v/v) Ficoll PM-400 were added to the RT reaction.
- Pre-heating step: RT primer, dNTPs and cell solution were mixed, everything was incubated at 55°C for 5 min and rapidly cooled down on ice before the remaining reagents were added.

6.2.2 Methanol fixation of tissue and cells

Methanol fixation of cells for single-cell experiments was carried out as previously described (Alles et al., 2017). In brief, cells harvested by trypsinization were twice washed with PBS, resuspended in PBS + Ribolock (1 U/µl) to a final concentration of $5x10^6$ cells/ml. Then, the 4-fold volume of ice-cold methanol was added dropwise while the cells were slowly vortexed. Methanol-fixed cells were either used directly or stored at -80°C. For usage, cells were pelleted (2500xg, 10 min, 4°C), rehydrated with PBS+ 0.01 % (w/v) BSA + Ribolock (1 U/µl) and washed once more with PBS+BSA+Ribolock before they were passed through a 35 µm cell strainer to get rid of clumps.

Methanol fixation of tissue was achieved by placing the 10 µl slices that were immobilized on Superfrost Plus microscope slides in cold methanol (15 min, -20°C). Next, the slices were rehydrated with PBS + Ribolock (1 U/µl) (3 washes).

6.2.3 Adapter ligation

Ligation of the adapter holding the second barcode is achieved with T4 DNA Ligase. To aid ligation a splint oligo is used that is first annealed to the adapter:

Reagent	Volume [µl]
splint oligo [100 µM]	1.0
adapter with 2^{nd} barcode [100 µM]	1.0
10x NEB CutSmart Buffer	0.5
water	2.5

This reaction mixture is incubated at 95°C for 5 min and then slowly cooled down to 20°C at a rate of 0.1°C/s. For the actual ligation reaction the following reaction setup is used:

Reagent	Volume [µl]
Cell Solution	3.00
10x NEB T4 DNA Ligation Buffer	1.00
T4 DNA Ligase (400 U/µl)	0.25
water	2.50

The ligation reaction is incubated at room temperature for 1 h, slowly shaking on an eppendorf MixMate (500 rpm).

6.2.4 Tissue digestion

For digestion of methanol-fixed and rehydrated tissue several proteases were tested for 30 min at 37°C each: Papain at 2 g/ml EBSS, TrypLE Express 1:1 in EBSS, Collagenase Type IV at 5 ml/ml in PBS/EBSS (1:1), and Proteinase K at 0.02 mg/ml in EBSS.

Finally, Papain at 2 g/ml in EBSS (37°C, 15 min and 37°C, 15 min, shaking at F9, 12 rpm on a neoLab Intelli-Mixer) were used in all experiments to digest coronal mouse brain slices.

6.2.5 Estimation of AMPure XP bead purification yields

For estimation of AMPure XP bead purification yields, 1 µg of RNA was reverse transcribed and split in two aliquots. One aliquot was kept (A), the other one (B) was mixed with either 40 µl of water, 40 µl of Dropseq Lysis buffer, 40 µl of water and Triton X-100 (final concentration 0.15 % v/v), or 40 µl of water and IGEPAL CA-630 (final concentration 0.25 % v/v) before the cDNA was purified with 90 µl AMPure XP beads according to the manufacturer's instructions. cDNA was eluted in 30 µl water, while the other aliquot (A) was brought to the same volume with water. qPCR was used to calculate the recovered amount of cDNA:

$$recovered\ cDNA\ [\%] = 2^{(C_t(A) - C_t(B))} \cdot 100\ \%$$

6.2.6 Design of CAD files

CAD files of the grid and its mounting device were created with Autodesk Fusion 360 (student license).

6.2.7 Production of a grid with a 3D printer

CAD files were loaded into Ultimaker Cura software to slice the models (.gcode file) for 3D printing with an ultimaker 2+. Ultimaker 2+ was set up with the manufacturer recommended print settings for the respective materials (extrusion speed, layer thickness minimal, extrusion temperature, plate heating etc.). To aid sticking of the first layer on the plate, glue from a conventional glue stick was applied. To ease release of the mounting device from the ultimaker 2+ plate, a scalpel was used.

Grids with low wall thickness (< 300 µm) were produced with a Projet HD 3000 at the Fraunhofer Institute for Applied Polymer Science in Potsdam Golm, with the help of Dr. Henry Memczak.

6.3 Cell Culture Methods

6.3.1 Culture and differentiation of SH-SY5Y

SH-SY5Y cells were cultured and differentiated as previously described (Ross et al., 1983). In summary, cells were grown in DMEM/F12 (Gibco) with 10 % FBS and 1x GlutaMax (Gibco). For differentiation, cells were plated at a density of $1x10^5$ cells/cm^2 and induced by addition of differentiation medium (Neurobasal medium, 1x GlutaMax, 1x B27, 10 µM ATRA) and harvested 4 or 8 days later. At day 4 post-induction, 5 µM AraC were added to prevent overgrowth of the culture with mitotic cells. The medium was exchanged every second day.

6.3.2 Culture and differentiation of NTERA2

NTERA-2 cells were cultured and differentiated as previously described (Lee and Andrews, 1986; Pleasure and Lee, 1993; Podrygajlo et al., 2009). In brief, $1x10^6$ cells/cm^2 were plated in bacteriological grade dishes in DMEM/F12 (Gibco), 10 % FBS and 1x GlutaMax (Gibco). After 2 days 10 µM ATRA was added and medium was renewed every two days until day 8. Embryoid bodies were then plated on a cell culture grade dish (same size as initial bacteriological dish) and grown as monolayer. At day 14 cells were trypsinized (5 min, room temperature, 0.05 % Trypsin/EDTA) and replated on a larger dish. At day 16 cells were trypsinized again and replated on a smaller dish while the media was additionally supplemented with 1 µM AraC. Finally, at day 23, cells were replated on PLL-coated dishes at a density of $4x10^5$ cells/cm^2 as neuronal culture.

6.3.3 Preparation and culture of primary neurons

Primary forebrain neurons were prepared from CD1 mouse embryos (E17.5-18.5) as previously described (Dotti et al., 1988; Kaech and Banker, 2006). They were maintained in Neurobasal-A medium with 2% B27 and 0.5 mM GlutaMAX-I (Gibco) at 37ºC and 5% CO2.

6.4 Computational Methods

6.4.1 Analysis of RNA sequencing samples

Sequencing data was converted to the fastq file format and demultiplexed using *bcl2fastq* and mapped to hg38 with STAR 2.4.1d or to mm10 with STAR 2.4.2a. Uniquely aligned reads were intersected with GENCODE v23 gene models (human) or GENCODE vM16 (mouse) and counted with HTseq (Anders et al., 2015) (human) or with GenomicRanges (Lawrence et al., 2013) and Rsubread (Liao et al., 2013).

6.4.2 Differential expression analysis

Differential expression changes were determined with DESeq2 (Love et al., 2014a). DESeq2 relies on the assumption that the majority of expressed genes is constant in between analyzed samples and comparatively little genes are changing in expression. Hence, for normalization DESeq2 finds scaling factors to normalize the gene expression distributions for different data sets. The principal component analysis was performed with *plotPCA* on regularized log transformed read counts.

6.4.3 GO term analysis

For GO term analysis, up- and downregulated genes were selected according to an expression cut-off (> 50 counts in scramble KD) and a significance cut-off for differential expression (human: adjusted p-value < 0.05; mouse: adjusted p-value < 0.001). Those selected genes were tested for overrepresentation of specific terms against all expressed genes (> 50 counts in scramble KD) with *goana* from the limma package (Ritchie et al., 2015). p-values are calculated with a Fisher's exact test and adjusted with the Benjamini-Hochberg method (Benjamini and Hochberg, 1995).

6.4.4 Identification of chimeric reads

Please refer to (Grosswendt et al., 2014) for a detailed description on how to identify chimeric reads. The dataset used for chimera analysis of mouse neocortex was published by Moore and colleagues (Moore et al., 2015).

6.4.5 Prediction of RCMs and calculation of H-score

The H-score is a probabilistic score which defines how likely an exon, or a group of exons will form a circRNA. H takes into account the length of RCMs and the distance between them. Importantly, H considers the competition of circRNA flanking RCMs with all other RCMs in pre-mRNA. RCMs are reverse complementary matches (sequences) located in circRNA flanking introns. RCMs form double-stranded structures across exons and promote circularization (Rybak-Wolf et al., 2015). Please refer to (Ivanov et al., 2015) for a detailed description.

6.4.6 Conservation analysis

PhyloP scores (Pollard et al., 2010) from the multiple alignment of 100 vertebrates were retrieved for hg19 using the UCSC table browser (Hinrichs et al., 2006). The mean was calculated for the following genomic regions (hg19): SLC45A4 exon 1 (circRNA exon) chr8:142,264,087-142,264,728; SLC45A4 exon 2 chr8:142,231,675-142,231,864; SLC45A4

exon 3 chr8:142,229,747-142,229,929; SLC45A4 exon 8 (CDS and 3'-UTR) chr8:142,220,870-142,221,800; SLC45A1 exon 1 chr1:8,377,887-8,378,248; SLC45A2 exon 1 chr5:33,984,302-33,984,835; SLC45A3 exon 1 chr1:205,633,612-205,634,014.

6.4.7 Reanalysis of single-cell data and marker gene determination

Single-cell data from mouse cortex E15.5 was taken from (Yuzwa et al., 2017) and downloaded from GEO (GSE107122, "GSE107122_E155_Combined_All_Cells_DGE.txt"). Seurat was used for the re-analysis of this dataset (Butler et al., 2018). Only non-mitochondrial genes that were expressed in ≥ 3 cells and cells that had ≥ 200 or ≤6000 genes expressed were kept. Data was normalized (*NormalizeData)*, variable genes were identified (*FindVariableGenes),* data was scaled (*ScaleData)* and principal components identified (*RunPCA)* and statistically significant principal components identified (*PCElbowPlot*, 1 to 16 were selected) with standard settings. Next, 14 cell clusters were identified (*FindClusters*, dims.use=1:16, resolution = 1.0), visualized with non-linear dimensional reduction (tSNE) as implemented in Seurat. Marker genes for these clusters were identified using the build in *FindAllMarkers* function and then individually evaluated for specific expression in the respective clusters. Based on these marker genes, the identity of the clusters was determined by literature research.

6.4.8 Inference of cell type proportion changes from single-cell and RNA-seq data

Marker genes for the different cell populations present in E15.5 mouse cortex were defined by reanalyzing single-cell data (Yuzwa et al., 2017) (GSE107122, "GSE107122_E155_Combined_All_Cells_DGE.txt"). Bseq-SC (Baron et al., 2016) was used to infer the changes in cell type proportions. An extensive tutorial for Bseq-SC can be found at https://github.com/shenorrLab/bseqsc.

6.5 Computational methods related to 3D-seq

All code for analysis of 3D-seq sequencing runs was written by Dr. Nikolaos Karaiskos, the latest version is accessible via https://github.com/nukappa/space-seq.

7 Results

7.1 Circular RNAs in the Mammalian brain are highly abundant, conserved and dynamically expressed

7.1.1 CircRNAs are abundant in the mouse and human central nervous systems

The detailed description of circRNA expression in a wide variety of neural samples allows for the discovery of many insights that later on potentially help to learn about the function of circRNAs. Knowledge of expression patterns, localization and spatiotemporal expression dynamics is necessary to identify those circular RNAs that are important for neuronal differentiation, synaptic plasticity and neuronal activity.

Therefore, total RNA libraries from different brain regions, primary neurons, synaptoneurosomes and cell lines over the course of differentiation were prepared (Rybak-Wolf et al., 2015). Additionally, data made available by the ENCODE consortium were analyzed (ENCODE Project Consortium, 2012), amounting in total to 28 samples (Table 3).

For detection of circRNAs a previously published approach (Memczak et al., 2013), that allows for *de novo* detection of circRNAs from total RNA sequencing data, was applied. CircRNAs are identified based on reads uniquely mapping to the head-to-tail splice junction of the circRNA, which is a splice junction only found in circular RNAs or in rare cases in *trans*-spliced transcripts. The head-to-tail junction spanning reads are used for quantification of a circRNA. To have a comparable measure for mRNAs, their quantification was limited to the splice sites as well. The maximum number of reads of the shared splice sites between linear and circRNA were counted (Figure 9 - A) to minimize effects on expression counts derived from other underlying sequence features, e.g. the length of surrounding introns (Rybak-Wolf et al., 2015). The cut-off for calling a circRNA expressed was set to two uniquely mapping, head-to-tail splice junction spanning reads.

Interestingly, the fraction of spliced reads falling on circRNA splice junctions differs between tissues, e.g. almost 5 % of spliced reads in human cortex can be assigned to circRNAs, while less than 1 % of spliced reads in liver originate from circRNAs (Figure 9 - B). This generalization of circular RNAs being highly expressed in neural tissues was validated by qPCR for circRIMS2, a highly expressed and conserved circRNA that is almost exclusively found in neural tissues of both mouse and human (Figure 9 - C).

Next, the number of reads falling onto the head-to-tail splice junction of a circRNA or on the adjacent linear splice junction of the linear RNAs was analyzed. It was revealed that linear RNAs are generally higher expressed than circular RNAs. However, circRNAs span a wide expression range and a considerable amount is expressed at more than 100 reads falling onto the head-to-tail junction: 565 (3.1 %) mouse circRNAs and even 5962 (8.4 %) human circRNAs (Figure 9 - D).

Additionally, circRNAs were frequently found to be the higher expressed isoform from a locus. In mouse post-natal hindbrain and midbrain, in more than 25 % of the genetic loci that form circRNAs, the circRNA is the higher expressed isoform (Table 3). In human occipital lobe and

temporal lobe still 11.0 % or 12.4 % of circRNA-forming genetic loci express the circRNA at a higher level than the mRNA.

Table 3: Overview of samples prepared and/or analyzed

organism	sample	# circRNAs detected	# circRNA > linear RNA	
mouse	cerebellum	2407	238	9.9 %
	pre-frontal cortex	1677	166	9.9 %
	hippocampus	1212	164	13.5 %
	olfactory bulb	1771	189	10.3 %
	P19 differentiation	2735	508	18.6 %
	undifferentiated, day 0	419	81	19.3 %
	day 2	363	74	20.4 %
	day 4	1128	153	13.6 %
	day 12	1210	117	9.7 %
	post-natal forebrain (ENCODE)	6614	1182	17.9 %
	post-natal hindbrain (ENCODE)	3382	1055	31.2 %
	post-natal midbrain (ENCODE)	3758	1033	27.5 %
	primary cortical neurons	5265	646	12.3 %
	day 1	1464	176	12.0 %
	day 7	1949	287	14.7 %
	day 14	2717	343	12.6 %
	day 21	2558	369	14.4 %
	synaptoneurosomes	1773	317	17.9 %
human	cerebellum (ENCODE)	21071	2289	10.9 %
	diencephalon (ENCODE)	24632	2430	9.9 %
	frontal cortex (ENCODE)	38983	4089	10.5 %
	occipital lobe (ENCODE)	31085	3410	11.0 %
	parietal lobe (ENCODE)	23303	2159	9.3 %
	temporal lobe (ENCODE)	21835	2718	12.4 %
	SH-SY5Y differentiation	4264	639	15.0 %
	undifferentiated, day 0	2166	254	11.7 %
	day 2	1822	193	10.6 %
	day 4	1300	174	13.4 %
	day 8	483	83	17.2 %

Next, we were interested to see from which genomic loci circular RNAs are formed preferentially. In contrast to mRNAs, circRNAs are spliced from a subset of genomic features of a genomic locus, e.g. a few exons. Therefore, we classified the circRNAs into circRNAs being produced from the 5'UTR, 3'UTR, coding sequence (CDS) or intronic regions of a gene or from ncRNAs, antisense of annotated genes or intergenic. Strikingly, most circRNAs were produced from the CDS of a gene, for both human (68.2 %) and mouse (55.0 %) (Figure 9 - E).

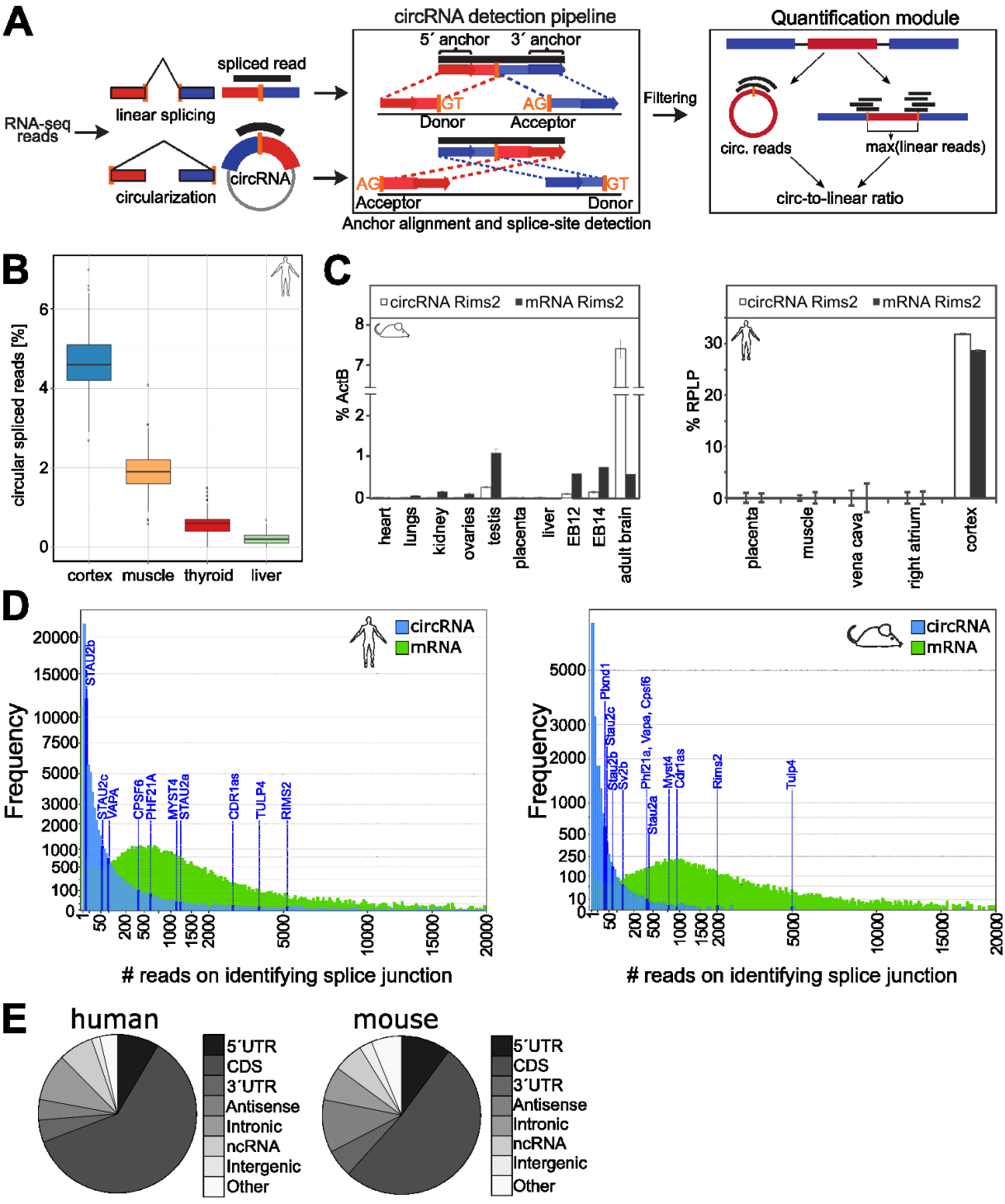


Figure 9: circRNAs are highly abundant in the mammalian brain.

(A) - circRNAs were detected as previously described (Memczak et al., 2013), and the analysis was extended by a comparison of circRNA and host transcript expression in deep sequencing data. **(B)** – Box plots of circular spliced reads in 4 human sequencing libraries, the highest fraction was assigned to circular splice junctions in cortex (blue), compared to muscle (yellow), thyroid gland (red), and liver (green). **(C)** - The conserved circRims2 is highly expressed in mouse adult brain (left panel; EB12/14, embryonic brain day 12/14) or human cortex (right panel). Data represent mean ± SD, n = 3 technical replicates. **(D)** - circRNAs detected in mouse and human brain samples span a broad expression range. However, mRNA expression is generally higher. mRNA – green, circRNA – blue. **(E)** – circRNAs stem mostly from the coding sequence of a gene (CDS).

7.1.2 CircRNAs are differentially expressed in the brain and are enriched at synaptoneurosomes

Since circRNAs were found to be predominantly expressed in neural tissue, like cortex, their spatial expression in the mouse brain was investigated. First, circRNA expression across different brain regions, e.g. cerebellum, hippocampus, olfactory bulb and prefrontal cortex, was compared by the number of normalized reads falling onto the head-to-tail junction. Strikingly, circRNA expression was found to be specific to each of these brain regions while a set of circRNAs exists in all brain regions (Figure 10 – A). Then, a number of circRNA candidates was selected and their differential spatial expression was confirmed independently by qPCR. Additionally, their expression was compared to their linear counterparts. In many instances the spatial expression of a circular RNA was independent from its linear mRNA counterpart, e.g. circRims2 is higher expressed in cerebellum than the mRNA, while the mRNA is more abundant in the striatum (Figure 10 – B). These observed expression differences for circRNAs in several mouse brain regions indicate a regulated way of expression for circRNAs. Additionally, the many examples of differential expression of circRNAs compared to their linear counterparts (Figure 10 – B) suggest an independent function of these circRNAs as well as supporting the hypothesis that circRNA expression is regulated independent of its linear counterpart.

Neurons are characterized by the long extensions they extrude from their soma, known as neurites or more specifically as axons and dendrites. Along these neurites synapses are found, structures that permit the signal transduction between two cells. The signal conducting and receiving parts of a synapse are known as pre- and post-synaptic compartments, the combination of both as synaptoneurosome. It was previously shown that certain mRNAs are actively transported to the synapses to facilitate for example local translation at the synapses (Schuman, 1999). Therefore, it was analyzed whether circRNAs could be localized to the synapse as well, possibly serving specialized functions in these cell compartments. Thus a well-established protocol to extract synaptoneurosomes by centrifugation from whole mouse brain, was used to separate those from whole brain extract and the cytoplasmic fraction (Gray and Whittaker, 1962; Huttner et al., 1983). Analysis of the subcellular localization of circRNAs by total RNA-seq of whole mouse brain extract, cytosolic fraction and synaptoneurosomes revealed their predominant enrichment at the synaptic regions, compared both to their linear counterparts and the whole brain extract (Figure 10 – C). Here, at least 5 uniquely mapping reads on the head-to-tail junction were required to consider a circRNA to increase the confidence in the data. To make sure that this observation was not based on inherently noisy low expression counts of the investigated molecules, the data was split up by circRNA expression thresholds (Figure 10 – D). Remarkably, the enrichment of circRNAs at the synaptoneurosome compared to either whole brain extract or cytoplasmic fraction was robust over all investigated expression thresholds. Again, we confirmed the previous observations from RNA-seq by independent low-throughput methods, in this case by qPCR (Figure 10 – E).

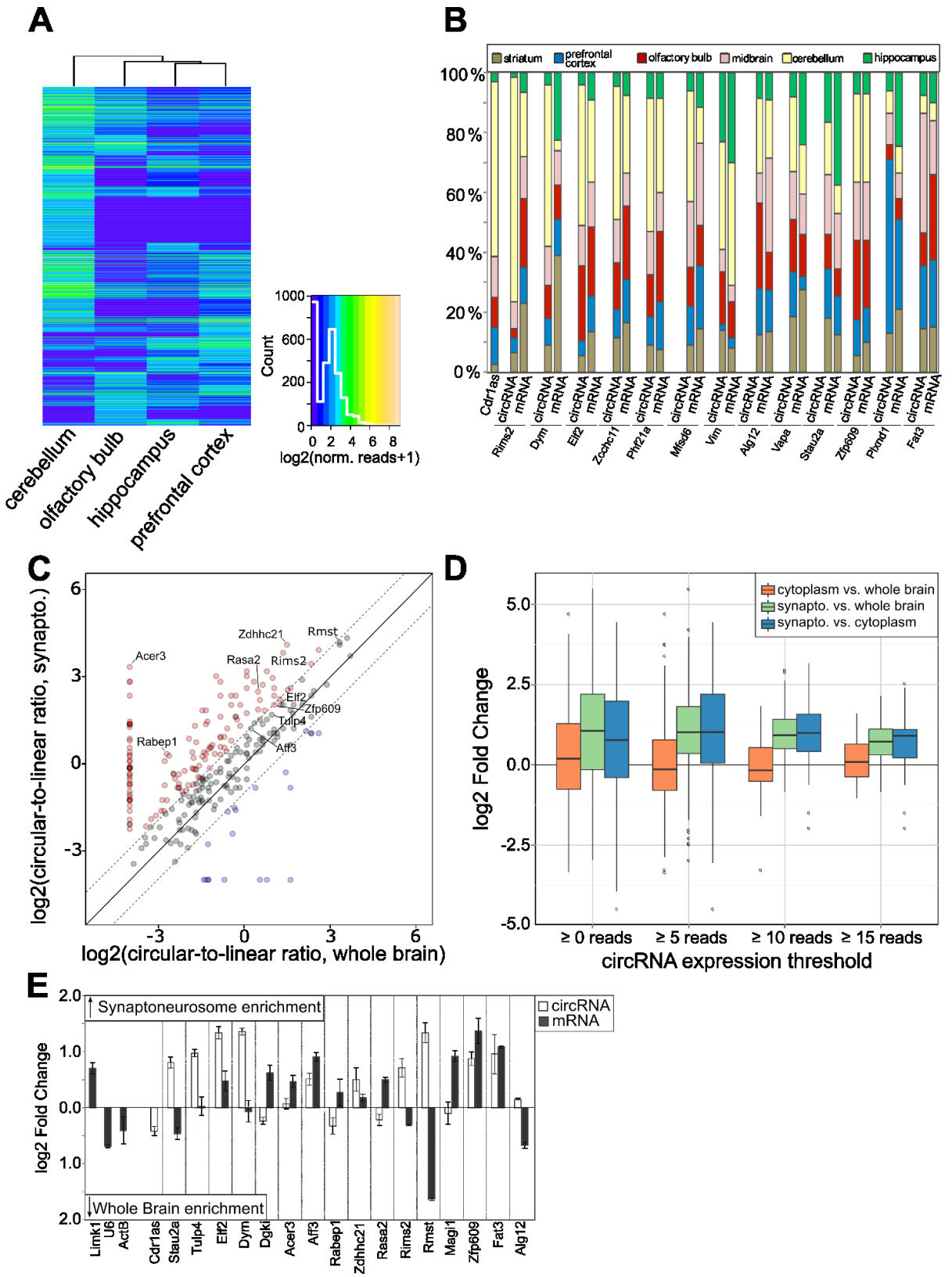


Figure 10: circRNAs are differently expressed in brain regions and neuronal compartments.
(A) – Clustering of mouse brain regions by normalized circRNA expression. **(B)** – qPCR validation of circRNA and mRNA expression in several mouse brain regions. **(C)** – circRNAs are enriched in synaptoneurosomes compared to whole brain extract, normalized by circular-to-linear ratio. Dashed line marks two-fold enrichment. **(D)** – CircRNA enrichment at synaptoneurosomes is robust over several circRNA expression cut-offs. **(E)** – Validation of circRNA enrichment at synaptoneurosomes by qPCR.

7.1.3 CircRNAs are differentially expressed during neuronal differentiation

For a functional characterization of particular circRNAs, established cell culture models are needed that allow for well-understood and controlled differentiation of progenitor cells towards more neuronal cells. For this purpose the human neuroblastoma cell line SH-SY5Y (Ross et al., 1983) as well as primary mouse forebrain neurons were analyzed for their circRNA expression throughout differentiation (Dotti et al., 1988; Kaech and Banker, 2006). In brief, SH-SY5Y cells were platted in conventional cell culture dishes and treated will all-trans retinoic acid. Primary neurons were prepared from hippocampi of embryonic mouse brains (E17.5 – E18.5) and grown on PLL-coated coverslips during maturation. Successful differentiation was validated by measuring neuronal marker genes, i.e. beta-III-tubulin, on RNA and protein level (data not shown).

The expression of circRNAs and mRNAs over the course of 4 days of differentiation was analyzed in human SH-SY5Y and primary mouse neurons by total RNA-seq. Comparison of circRNA expression in undifferentiated (D1) to differentiated (D21) primary neurons revealed that almost 2000 circRNAs were upregulated while only 800 were downregulated (Figure 11 – A). Interestingly, many circRNAs were not expressed at day 1 but are only detected later on. This points towards a general trend of circRNAs being upregulated over the course of neuronal differentiation.

The same analysis for SH-SY5Y cells showed a similar however weaker trend, with 494 circRNAs being upregulated when comparing undifferentiated cells to day 4 of differentiation. At the same time 539 circRNAs are downregulated in SH-SY5Y cells. It should be kept in mind, that so called undifferentiated SH-SY5Y cells are already at a neuroblast stage, which is an already advanced step in neurogenesis that permits cells to become mature neurons (Kovalevich and Langford, 2013). Additionally, these numbers reflect global changes and do not consider slight but consistent upregulation of circRNAs overall which will be explored in the next section.

Again, several candidates from both systems were selected from total RNA-seq results and validated by qPCR (Figure 11 – B and D). Often the same expression trend for a circRNA was identified in primary neurons as well as SH-SY5Y cells, e.g. CDR1as, CPSF6 and RTN4 are consistently upregulated, despite the fact that those differentiation systems are derived from different organisms.

Analyzing the expression of all circular RNAs over the course of differentiation individually revealed the expression dynamics for this molecule class in both SH-SY5Y cells and primary neurons
(Figure 11 – C and F). Most circular RNAs peak at a certain step of differentiation, while only few remain constant throughout differentiation. Many circRNAs are expressed only during intermediary steps of differentiation: e.g. some were only detectable at D2 for SH-SY5Y cells, at an early commitment point to the neuronal lineage, or at D7/D14 for primary neurons, when they are maturing. These intriguing expression dynamics are pointing at a tightly regulated expression of circular RNAs during the differentiation from progenitor cells towards neurons.

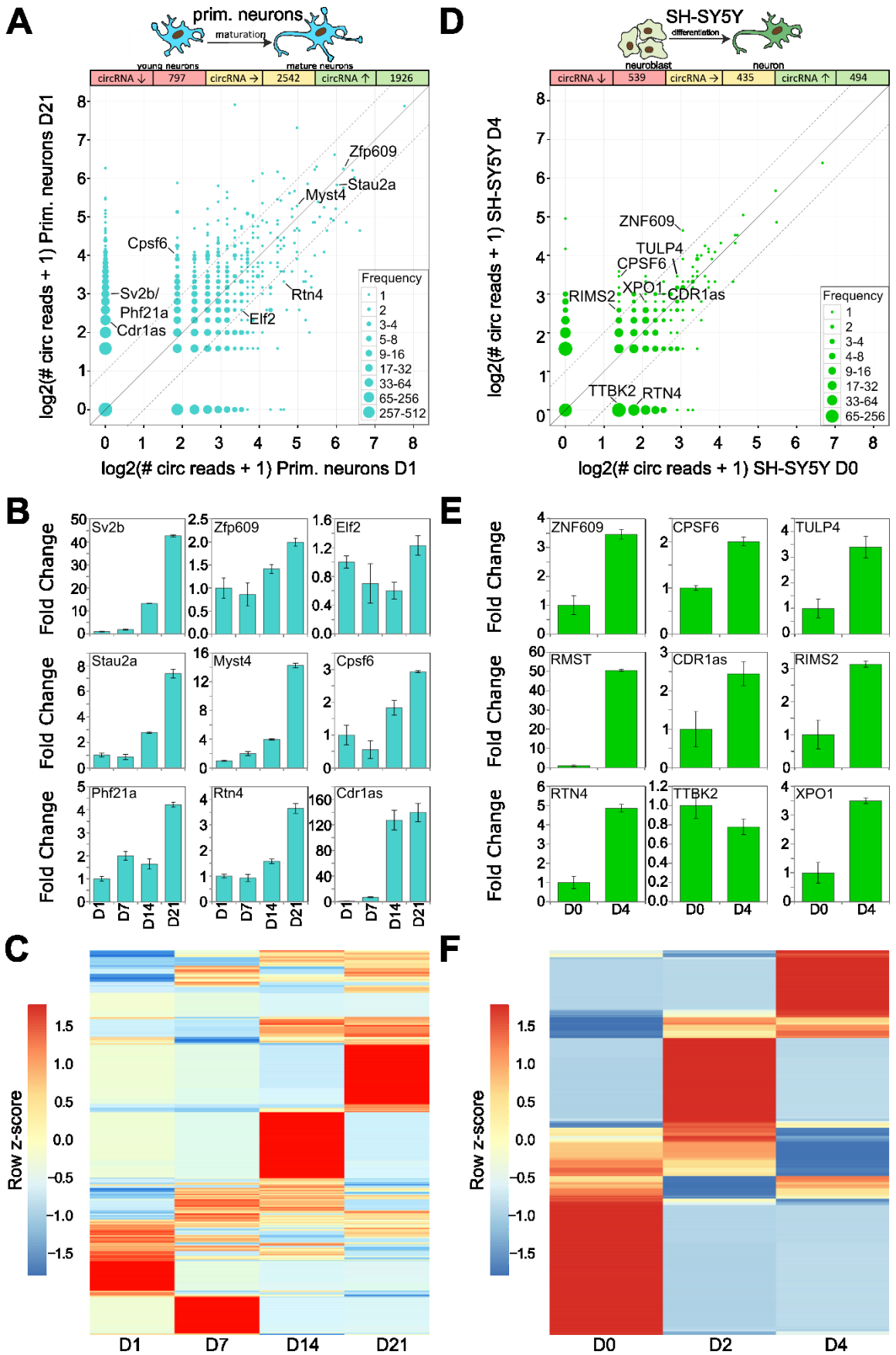


Figure 11: circRNAs are differentially expressed during neuronal differentiation.
(A) – circRNA expression comparison between D1 and D21 of primary neuron maturation. **(B)** – qPCR validation of primary neurons circRNA expression changes detected previously by RNA-seq. **(C)** – Heatmap of primary neuron circRNA expression changes. Each row corresponds to a circRNA, coloring is scaled to rows. **(D)** - circRNA expression comparison between D0 and D4 of SH-SY5Y differentiation. **(E)** – qPCR validation of SH-SY5Y circRNA expression changes detected previously by RNA-seq. **(F)** – Heatmap of SH-SY5Y circRNA expression changes. Each row corresponds to a circRNA, coloring is scaled to rows.

7.1.4 CircRNA and mRNA changes during differentiation differ

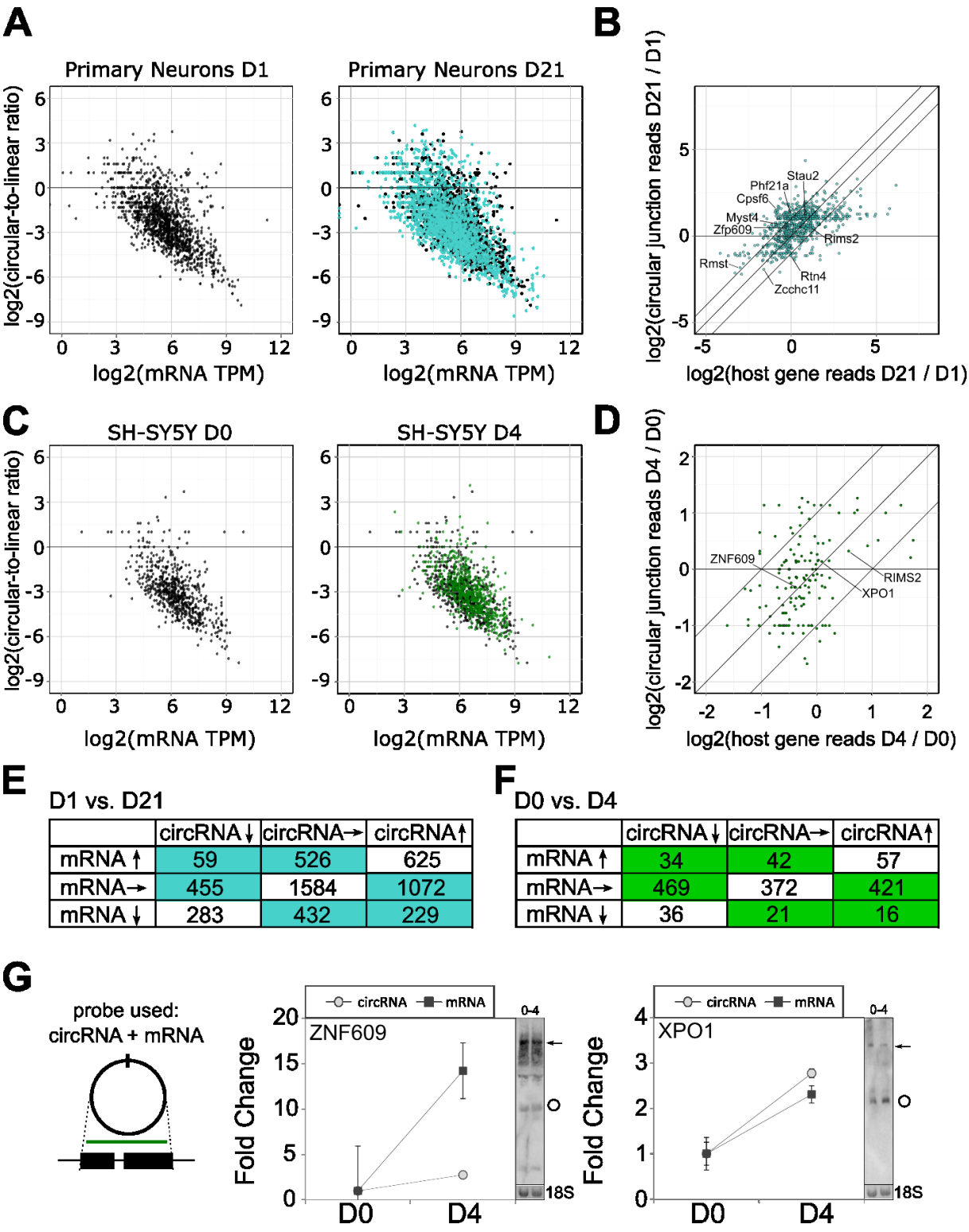
Considering, that circRNAs are alternative splicing events from a gene, it is of great importance to compare their expression changes with its linear splice isoform. mRNA expression levels are often measured in transcripts per kilobase million (TPM) (Wagner et al., 2012) correcting the expression inferred from RNA-seq count data for sequencing depth and gene length. Since circular RNAs are only identifiable by their unique head-to-tail junction a similar measure is not easily derived for circular RNAs.

So, to investigate expression changes for circular RNAs during differentiation in respect to its linear counterpart, we made use of the comparison circular-to-linear ratio vs. mRNA TPM. Clearly, these measures are negatively correlated meaning that whenever a circular RNA is expressed from a genetic locus, the mRNA is not. This points to a mode were either a circular RNA or a linear RNA is produced from a genetic locus, indicating competition between the production of both possible isoforms. This observation was made in primary cortical neurons as well as in human SH-SY5Y cells in both undifferentiated and differentiated states (Figure 12 – A and C). This finding also argues against the idea that circular RNAs are a mere by-product of linear splicing, in which case no correlation between circular-to-linear ratio and mRNA expression would be expected. Furthermore, for SH-SY5Y cells an overall increase of the circular-to-linear ratio is visible, supporting the fact that circular RNA expression is in general higher in differentiated cells than in undifferentiated ones.

Comparing reads mapping to the head-to-tail junction (circular RNAs) or to circRNA surrounding splice sites (mRNA) revealed that the overall expression of mRNAs and circRNAs is correlated, also during differentiation (Figure 12 – B and D). However, there are many exceptions found in both mouse primary neurons and human SH-SY5Y cells, in which the circRNA expression level is independently changed from the mRNA's one (Figure 12 – B and D).

In summary, the expression changes of a circRNA and its corresponding mRNA are often independent from each other (Figure 12 – B and D). Counting the number of events, revealed that in primary neurons 1301 circRNAs were more than two-fold upregulated at day 21 of differentiation compared to undifferentiated cells, while the mRNA remained stable (Figure 12 – E). Only 625 upregulated circRNAs had linear counterparts that exhibited the same expression trend. A similar observation was made in human SH-SY5Y cells: 437 circRNAs showed increased levels in differentiated cells and that independent of their linear counterpart. At the same time only 57 circRNAs were upregulated together with their mRNA (Figure 12 – F).

Several circRNA expression trends were validated by northern blotting. Northern blot allows amplification-free quantification of RNA levels (Figure 12 – G). For detection DIG-labeled RNA probes were used that were designed to detect circular and linear RNA at the same time.



E D1 vs. D21

	circRNA↓	circRNA→	circRNA↑
mRNA↑	59	526	625
mRNA→	455	1584	1072
mRNA↓	283	432	229

F D0 vs. D4

	circRNA↓	circRNA→	circRNA↑
mRNA↑	34	42	57
mRNA→	469	372	421
mRNA↓	36	21	16

Figure 12: circRNA and mRNA expression changes differ during differentiation.
(A) - Relative quantification of circRNA expression during differentiation of primary neurons, circular to linear ratio vs. mRNA TPM. **(B)** – circRNA expression is positively correlated with mRNA expression in undifferentiated vs. differentiated primary neurons. **(C)** - Relative quantification of circRNA expression during differentiation of SH-SY5Y cells, circular to linear ratio vs. mRNA TPM. **(D)** - circRNA expression is positively correlated with mRNA expression in undifferentiated vs. differentiated SH-SY5Y cells. **(E)** – Summary of expression changes for primary neurons. **(F)** – Summary of expression changes for SH-SY5Y cells. **(G)** – Validation of expression changes by Northern blot for SH-SY5Y cells, ZNF609 and XPO1. Circle = circular RNA, arrow – linear RNA.

7.1.5 Circular RNAs are often conserved

Tools to predict circRNAs became available only recently, hence little basic information about circRNAs is available. Thus, after having determined physiological human and mouse circRNA expression levels, the conservation of circRNAs between mouse and human was analyzed. First, using the liftOver tool (Hinrichs et al., 2006) orthologous circRNAs were identified in human and mouse. Then these circRNAs were assigned to different categories based on the conserved genomic position of utilized splice sites in these two species (Figure 13 – A). A splice site was only considered conserved if it did not vary by more than 2 nt in position. Thereby even small shifts in splice site position over 100 million years of evolution from mouse to human were excluded. All mouse circRNAs identified were compared to all identified human circRNAs, which revealed that 28.5 % of mouse circRNAs have a homolog in human (4522 out of 15849 considered mouse circRNAs). Another 28.6 % of mouse circRNAs had at least one of their splice sites in a homologous position in human.

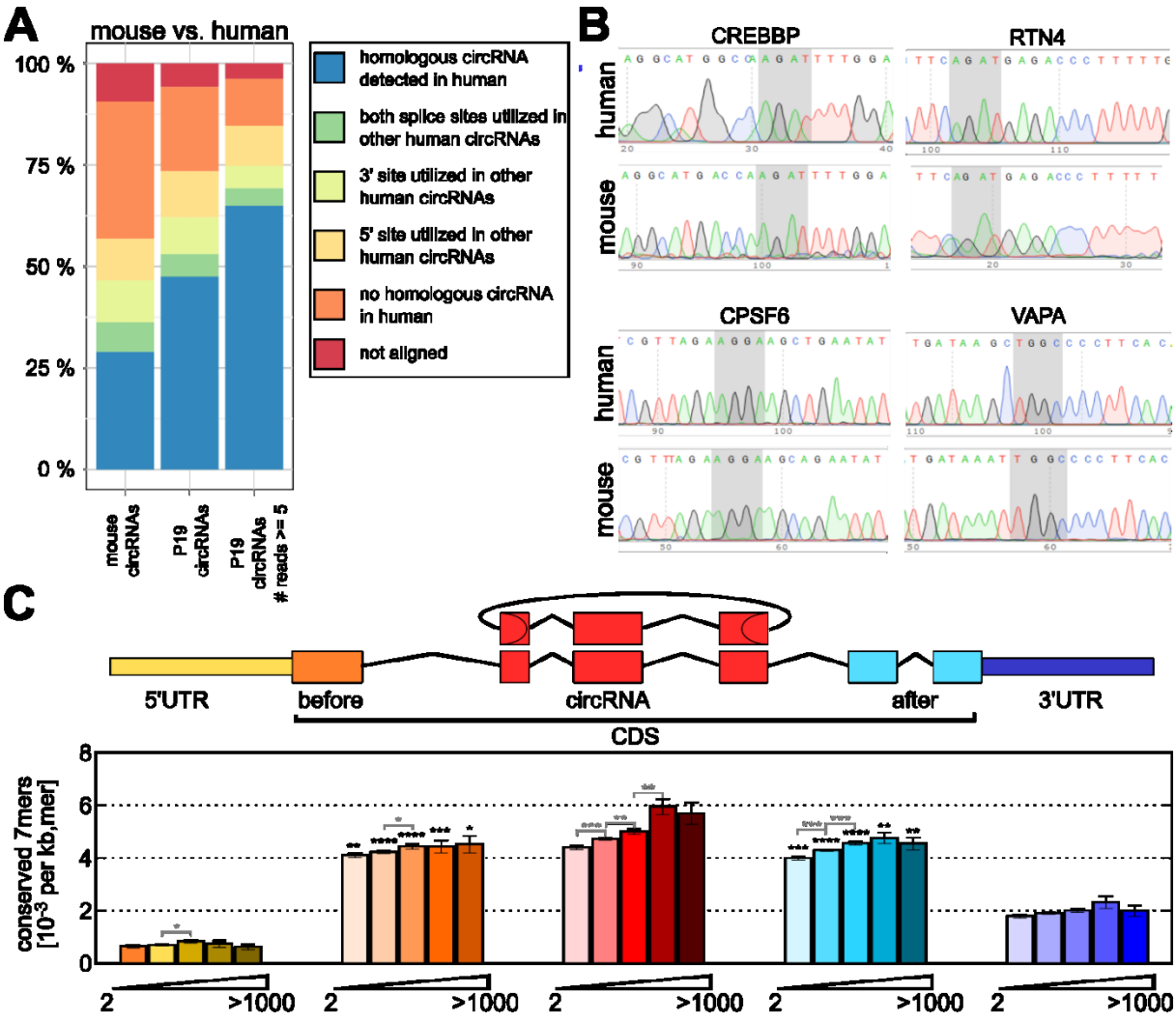


Figure 13: Conservation of circRNAs

(A) - Conservation analysis between human and mouse circRNAs using liftOver to identify homologous splice sites. Conservation levels are higher, if an expression cut-off (≥ 5 reads) is applied. **(B)** – Validation of several candidate circRNAs by Sanger sequencing confirming their homologous splice sites. **(C)** – Definition of analyzed segments (5'UTR, CDS – before, circRNA, after, and 3'UTR) and density of conserved 7-mers (total number of occurrences per total spliced length) up to chicken. Ramp underneath indicates expression levels: 2-10. 10-100, 100-500, 500-1000 and > 1000 head-to-tail spliced reads. Error bars represent SD based on random subsamples, asterisks: *p < 5%, **p < 1%, ***p < 0.1%, and ****p < 0.01%.

These numbers remained comparable when only circRNAs detected in a neuronal mouse cell line (P19) were examined for their conservation with human. However, the numbers increased dramatically when an additional expression cut-off of ≥ 5 reads on the head-to-tail junction for the P19 circRNAs was applied.

A number of conserved circRNA candidates was selected and the conservation of their head-to-tail junction was confirmed by Sanger sequencing (Figure 13 – B). Circularity was validated independently by RNase R treatment (data not shown), *i.e.* an exonuclease that specifically degrades linear transcripts (Suzuki et al., 2006).

Conserved circularization indicates selective pressure that preserves conditions in which the splicing machinery maintains to circularize specific exons throughout evolution. This could potentially be caused by independent functions of circRNAs that ought to be kept or simply because the underlying CDS is under selective pressure to keep the protein coding sequence. Therefore, it was examined if those parts of a CDS that give rise to a circular RNA were more conserved than those parts of the same CDS that are not circularized but are coding for a protein (Figure 13 – C). The analysis was limited to circRNAs coming from coding sequences, as this is the majority of circRNAs (comp. Figure 9 – E). For this purpose, conserved 7-mers (human up to chicken, 300 million years of evolution) per total spliced length were counted in the 5'UTR, coding sequence upstream and downstream of the circRNA, the circRNA itself and in the 3'UTR. Strikingly, those segments being used for circularization were enriched in conserved 7mers compared to any other gene segment. Additionally, this effect increased with higher expression of a circRNA. However, interpretation of this analysis is not straightforward: conservation is not homogenous along the CDS. More highly expressed circRNAs are located closer to the transcription start of the host gene and, more highly expressed circRNAs are flanked by longer introns which also correlates with higher conservation levels of exons (compare (Rybak-Wolf et al., 2015)).

7.1.6 Circular RNA expression is regulated by ADAR

ADAR1 is an A-to-I RNA-editing enzyme that has been implicated in the regulation of circRNA expression (Ivanov et al., 2015; Liang and Wilusz, 2014). It was shown that ADAR1 can hyper-edit reverse complementary sequences found in introns surrounding circRNAs. This editing activity would result in decreased pairing of the reverse complementary sequences, therefore ADAR1 negatively regulates circRNA expression. Additionally, it is known that ADAR1 is decreasing in expression over the course of human embryonic stem cells towards neuronal cells (Osenberg et al., 2010). Which would correlate with the above mentioned circRNA expression increase during differentiation.

Hence, the influence of ADAR1 on circRNA expression in the investigated neural samples was determined. First, the H-score that is used to identify reverse complementary sequence matches (RCM) was computed in order to predict circularization probability (Figure 14 – A). This revealed that especially in the introns surrounding conserved circRNAs extensive RCMs can be found. Often RCMs are formed by repetitive sequence elements that rapidly evolve, e.g. the big class of *Alu* elements which is only found in primates (Schmid, 1996). Consequentially, the H-scores for human samples are generally higher than for mouse samples.

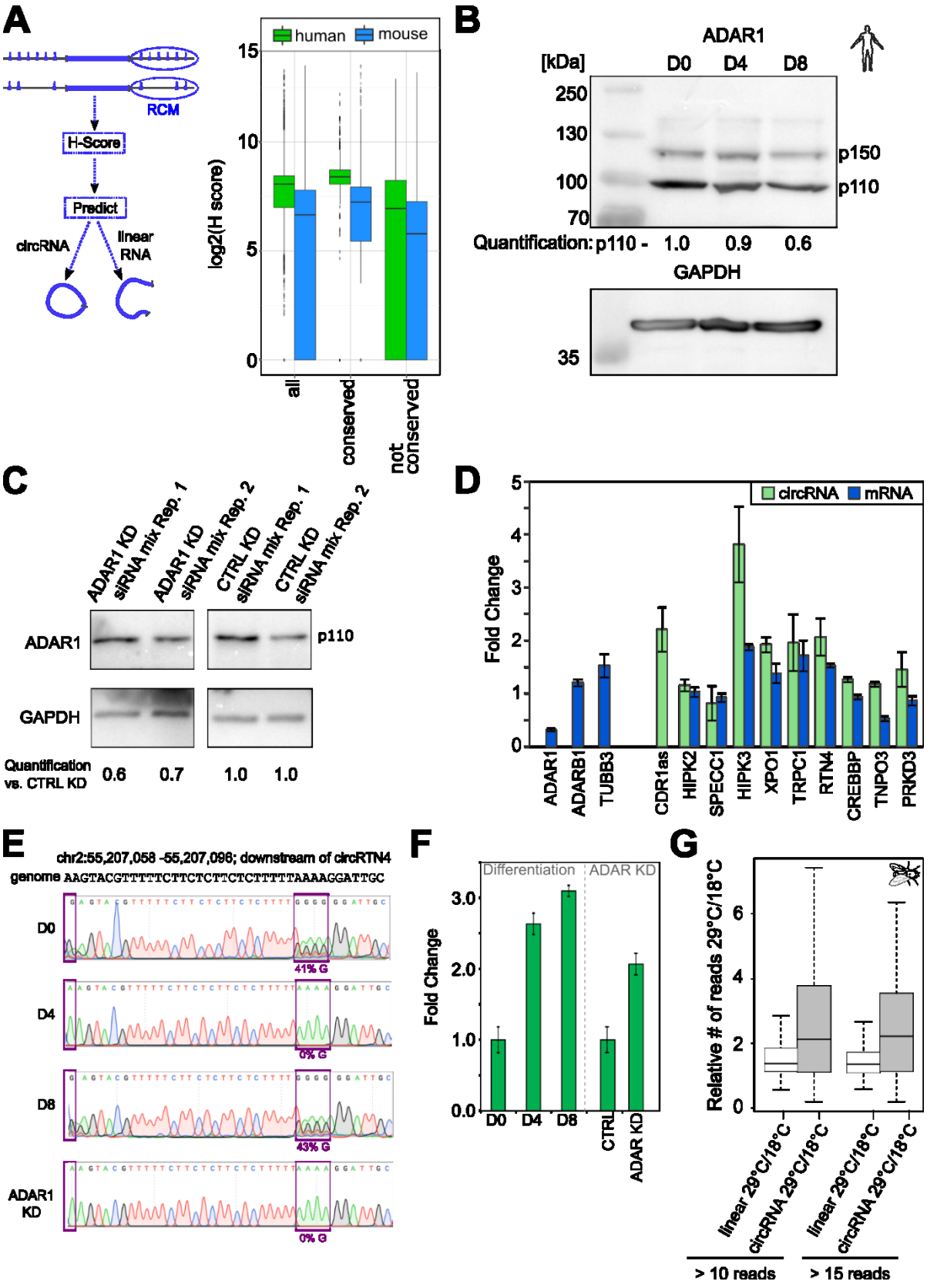


Figure 14: circRNA expression is regulated by ADAR

(A) – Concept of H-score calculation. H-scores for introns surrounding circRNAs in human and mouse, split up in conserved and not conserved circRNAs. **(B)** – ADAR1 expression during SH-SY5Y differentiation. Quantification for p110. **(C)** – Efficiency of ADAR1 KD in SH-SY5Y. **(D)** – circRNA and mRNA expression changes as determined by qPCR after ADAR1 KD in SH-SY5Y cells, compared to mock KD. **(E)** – Editing of region close to circRTN4 by ADAR1 during differentiation and after ADAR1 KD. **(F)** – Expression changes of circRTN4 during differentiation of SH-SY5Y cells and after ADAR1 KD. **(G)** – Global changes in circRNA expression in Drosophila at different temperatures that regulate dAdar activity (not active at 29°C). Boxes are limited by the first quartile and the third quartile, median is indicated as a horizontal bar, while whiskers span 1.5 x inter-quartile range from their hinges.

Since circRNAs were already observed to vary throughout neuronal differentiation the expression of ADAR1 over the course of SH-SY5Y differentiation was examined. A decrease of 40 % in protein expression after 8 days of differentiation was found (Figure 14 – B, p110) being in good agreement with the previously observed upregulation of circRNAs during differentiation. Knockdown of ADAR1 in SH-SY5Y cells (Figure 14 – C) depleted the protein by 30-40%. This knockdown caused striking changes in gene expression: almost all tested circRNAs were elevated in expression and much more so than their corresponding mRNA (Figure 14 – D).

Additionally, a region downstream of circRTN4 that is edited by ADAR1 on transcriptome level at certain points of differentiation was identified by computational analysis (Levanon et al., 2004)
(Figure 14 – E). As already mentioned, editing would result in downregulation of circRNA expression. At day 4 of differentiation the locus was not edited, therefore circRNA expression was released. Indeed, qPCR revealed the expected increase in circRNA expression (Figure 14 – F). At day 8 of differentiation the locus was edited again, which should prevent further elevation of expression, which was again confirmed by qPCR (Figure 14 – F). Furthermore, knockdown of ADAR1 would prevent editing of this site and cause the upregulation of circRTN4 to a similar extent as on day 4 of differentiation, which was again revealed by qPCR (Figure 14 – F).

To investigate whether this is a global and conserved phenomenon a previously published *Drosophila melanogaster* dataset was reanalyzed. Editing by dAdar, the *Drosophila melanogaster* homolog of ADAR1, is temperature-sensitive (Rieder et al., 2015; Savva et al., 2012) and therefore not active at higher temperatures (29°C). Comparing the relative number of reads on the head-to-tail splice junction to identify circRNAs or the adjacent splice junction to count linear RNAs revealed that only circRNA expression was increased at high temperatures, *i.e.* when dADAR is not active anymore. This is true for both investigated expression cut-offs, requiring circRNAs to have at least 10 reads on the head-to-tail junction (137 circRNAs) or 15 reads(78 circRNAs) (Figure 14 – G), and highlights the importance of circRNA expression regulation by ADAR as it is conserved even in *Drosophila melanogaster.*

7.2 CircSLC45A4 is required to keep neuronal cells in a progenitor state in cell culture systems and in the mammalian brain

7.2.1 CircSLC45A4 is conserved and highly expressed in embryonic frontal cortex

As mentioned above, circRNAs are abundantly expressed in neural samples, are conserved, are predominantly localized to the synaptoneurosomes, and are expressed in a spatiotemporally controlled manner (Rybak-Wolf et al., 2015). This collection of neural circular RNAs includes a variety of interesting molecules that are potentially important in neuronal differentiation. One of those being circSLC45A4, a circular RNA that is produced from an alternative first exon of SLC45A4 (Figure 15 - A). Interestingly, this circRNA is expressed in a dynamic spatial pattern in different brain regions. SLC45A4 is coding for a proton-dependent sugar transporter that can import glucose, fructose and sucrose, and was shown to be important in spermatozoa nutrition (Vitavska and Wieczorek, 2017). However, the function of circSLC45A4 remains elusive as it has never been studied in any biological context.

To survey circRNA expression in development, a total RNA-seq dataset of human embryonic frontal cortex at gestational week (WG) 22 was analyzed (collected by the ENCODE consortium (ENCODE Project Consortium, 2012)). Surprisingly, this dataset revealed that circSLC45A4, and not its linear counterpart, is the main RNA isoform transcribed from its genomic locus (Figure 16 – A). And it is overall one of the highest expressed circular RNAs in human cortex at this developmental time point. With 460 total reads spanning the head-to-tail junction circSLC45A4 ranks at 295th position in terms of circRNA expression in human fetal cortex at WG 22 of 38983 detected circRNAs with at least 2 unique head-to-tail junction spanning reads (Figure 15 - B), placing it in the top 1 % of expressed circRNAs. These 460 junction spanning reads indicate a very high expression of circSLC45A4 in this dataset. For comparison, amongst all circRNA-producing genes the median of all circRNA surrounding linear splice-junction spanning reads, or the median splice junction count for mRNAs, is 85. Considering both, linear splice junction spanning reads from mRNAs as well as circRNA splice junction spanning reads, circSLC45A4 would still be within the 99.5 percentile according to its expression. By the circular-to-linear ratio (refer to chapter 7.1.1) circSLC45A4 is ranked 6th of 36408 circRNAs (top 0.02 %), when only circRNAs are taken into account whose linear counterpart was at least detected with 2 unique splice-junction reads as well (Figure 15 - C). At WG 22, neurogenesis in the cortex is almost completed, while neuronal migration and maturation are just beginning (Stagni et al., 2015).

This ENCODE dataset also revealed a previously unannotated upstream exon in human that is necessary to facilitate the formation of circSLC45A4 by a splicing reaction. Intriguingly, not only is this upstream exon present in human and mouse (Figure 15 - D), but the mouse upstream exon is exactly duplicated in its length (420 nt instead of 210 nt in human) with the splice sites being conserved.

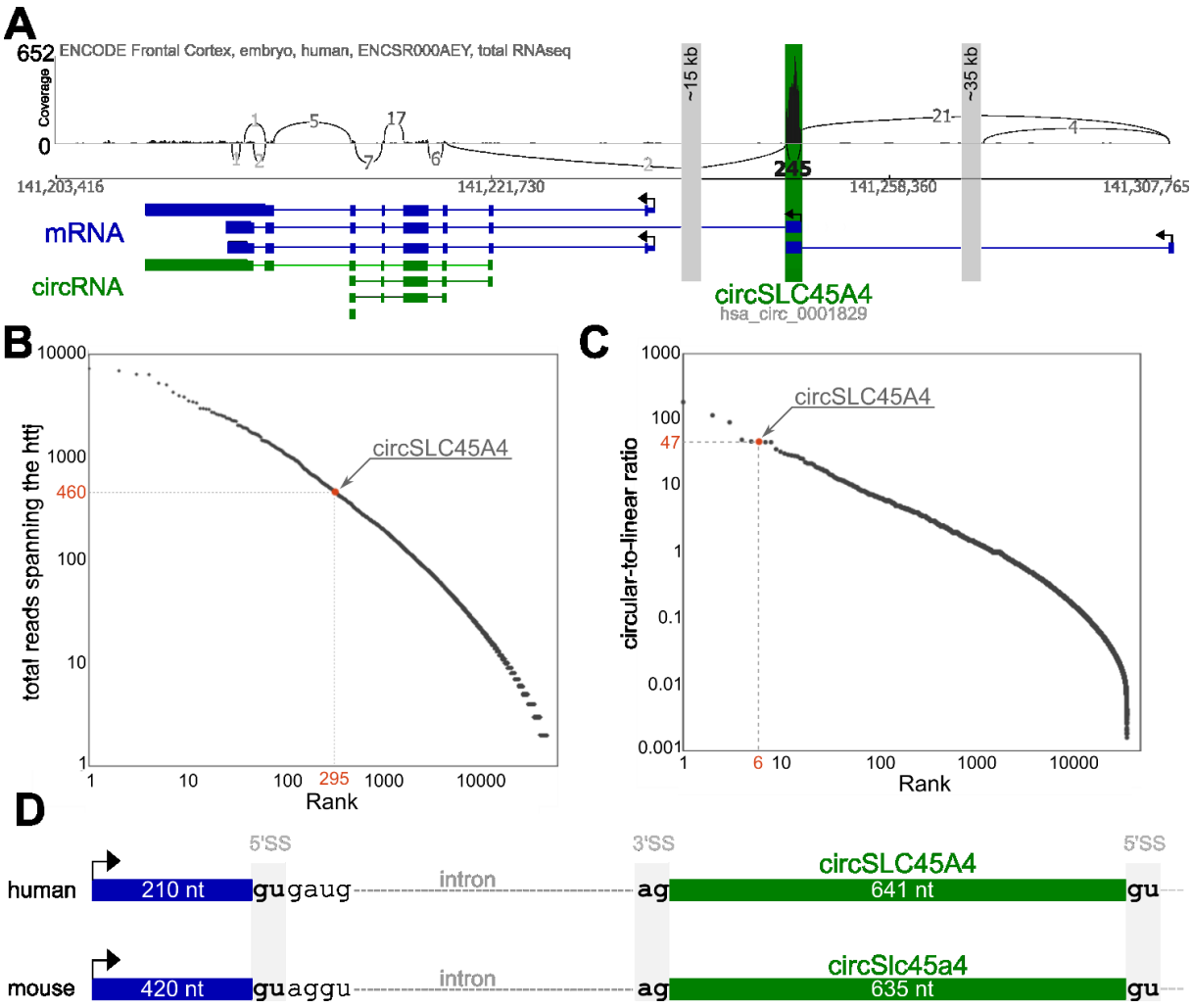


Figure 15: circSLC45A4 is a conserved circular RNA that is highly expressed in human embryonic frontal cortex. **(A)** – Genetic locus of circSLC45A4 in human. mRNA isoforms are shown in blue, circRNA isoforms in green. Coverage track shows total RNA sequencing of fetal human cortex, WG 22. Numbers on connecting lines indicate the number of reads supporting the splice junction. The location of circSLC45A4 (has_circ_0001829) is indicated. **(B)** – Expression rank of circSLC45A4 in fetal human cortex, ranked by total reads spanning the head-to-tail splice junction (295th, top 1 %). **(C)** – Expression rank of circSLC45A4 in fetal human cortex, ranked by circular-to-linear ratio (6th, top 0.02 %). **(D)** – Splice sites that facilitate production of circSLC45A4 in human and mouse.

The phyloP score (Pollard et al., 2010) computes conservation p-values based on an alignment and a model of neutral evolution. Using the phyloP scores that are made available through the UCSC genome browser based on the alignment of 100 vertebrates, the conservation of circSLC45A4 was assessed. This showed that the circRNA-forming exon of SLC45A4 is significantly more conserved than the 8th exon of this gene, containing CDS and 3'-UTR, but that it is also significantly more conserved than the first exons of genes from the same gene family (SLC45Ax) (Figure 16 – A). This locus is exhibiting unusually strong sequence conservation especially in those parts of the circular RNA that are overlapping the 5'-UTR of SLC45A4. Overall, circSLC45A4 is conserved on nucleotide level between human and *Xenopus tropicalis* (Figure 16 – B). Only 0.94 % (6 nt out of 641 nt) of the human to *X. tropicalis* alignment in that region belong to a gap. In contrast, the 3'-UTR and intronic regions of this gene are much less conserved (Figure 16 – C). Already in orangutan 4.75 % of this region alignment result in gaps, in *X. tropicalis* only 43 % of this region show conservation with human. This may suggest that the circRNA containing part of SLC45A4 is under a stronger selection pressure than other regions of this gene.

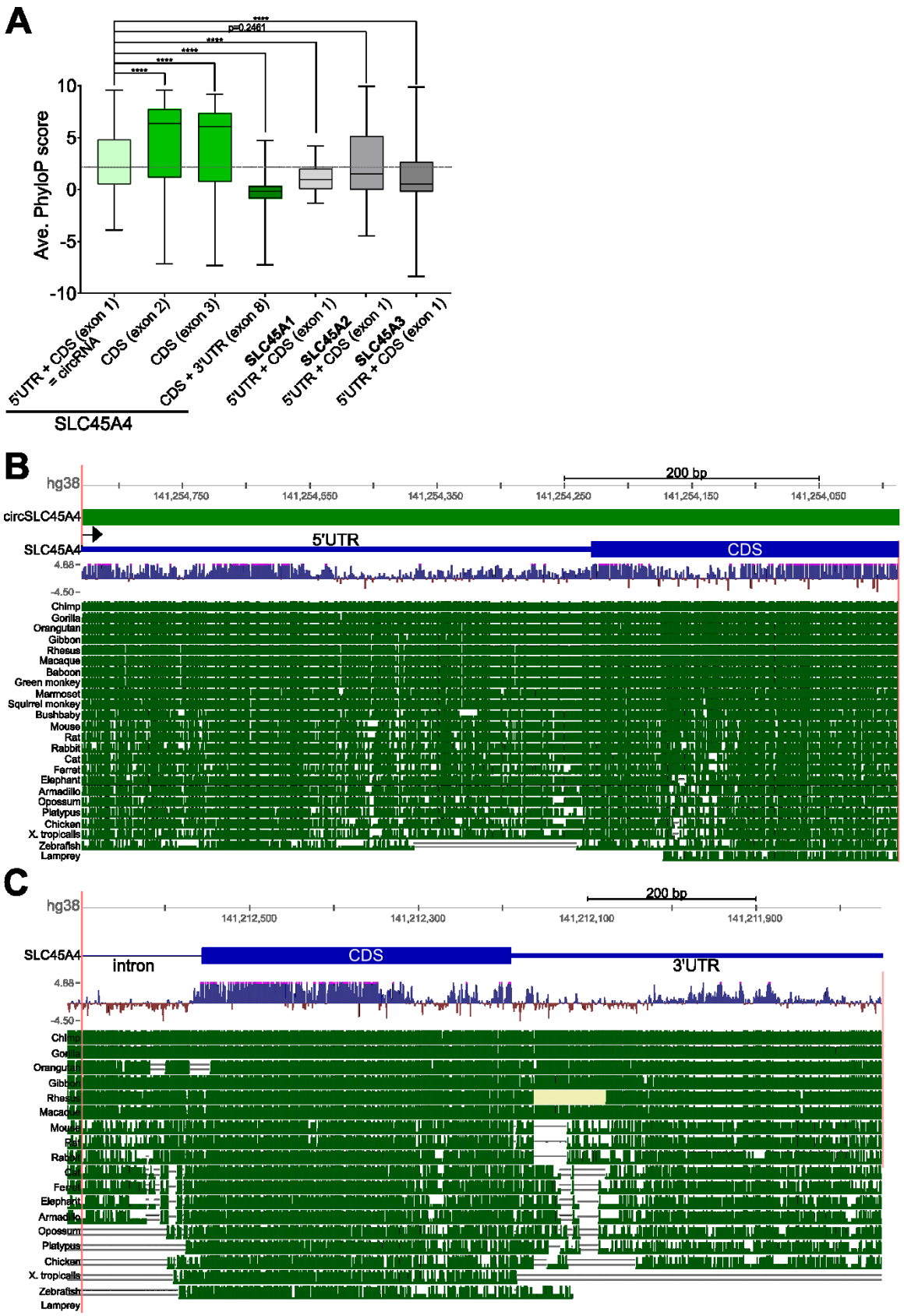


Figure 16: The genomic region of circSLC45A4 is a more conserved than comparable regions.
(A) – Average PhyloP scores of multiple alignments of 100 vertebra, SLC45A4 exon1 (641 nt, mean = 2.86), 2 (190 nt, mean = 5.04), 3 (183 nt, mean = 4.52), and 8 (931 nt, mean = -0.25), SLC45A1 exon 1 (103 nt, mean = 1.13), SLC45A2 exon 1 (534 nt, mean = 2.64), SLC45A3 exon 1 (403 nt; mean = 1.73). T-test, p-value < 0.0001 is ****. **(B)** –Genome Browser conservation tracks visualizing the exceptional conservation of circSLC45A4 compared to the 3'UTR and CDS of the same gene, SLC45A4 **(C)**.

7.2.2 CircSLC45A4 is localized in the cytoplasm and dynamically expressed during the neuronal differentiation of SH-SY5Y and NTERA2 cells

To investigate the functionality of circSLC45A4 suggested by the sequence conservation analysis, a model system for human neuronal differentiation was employed. SH-SY5Y cells are a well-established and widely used human cell line to study neuronal differentiation (Kovalevich and Langford, 2013; Ross et al., 1983). Using two-color single molecule FISH (Raj and Tyagi, 2010), an average of 3.7 and 8.8 molecules of circSLC45A4 and mRNA per SH-SY5Y cell were detected, respectively (Figure 17 – A). Fractionation experiments revealed that circSLC45A5 was exclusively found in the cytoplasm (93.9 % of total circRNA), while the mRNA was distributed throughout the cells: only half of the mRNA SLC45A4 RNA resided in the cytoplasm (56.4 %), while substantial amounts were localized to the nucleus (17.4 %) and to the mitochondria (26.2 %) (Figure 17 – B). The successful fractionation of SH-SY5Y cells was validated with ACTB and GAPDH as cytoplasmic markers, mitochondrial CO2 (MT-CO2) as mitochondrial marker, and NEAT1 as nuclear marker.

Next, the expression of circSLC45A4 and its corresponding mRNA over the course of neuronal differentiation in SH-SY5Y cells was analyzed. SH-SY5Y cells can be differentiated from neuroblasts to neurons by induction with all-trans retinoic acid (ATRA) (Påhlman et al., 1984; Ross et al., 1983). Here, it was revealed that the expression of circSLC45A4 peaks at the time of neuronal lineage commitment or birth of young neurons (Figure 17 – C). This coincides with elevated expression of beta-III-tubulin, a marker gene for the generation of immature postmitotic neurons (Menezes and Luskin, 1994). mRNA SLC45A4 is also induced at that time point, but less strong (1.4-fold compared to 1.8-fold induction of circSLC45A4).

A more complex human neuronal differentiation system is the teratocarcinoma-derived NTERA2 cell line that resembles *in vivo* differentiation of human neurons much closer (Lee and Andrews, 1986; Pleasure and Lee, 1993). Neuronal differentiation in these cells is induced via addition of all-trans retinoic acid. Upon this induction they go through an embryoid body (EB) stage. EBs are three--dimensional aggregates of pluripotent stem cells. From this stage it takes 12 days until maturing neurons are formed. circSLC45A4 was induced at a late stage of embryoid body formation, which coincides with the expression of PAX6, that in turn marks the neuroectoderm specification (Zhang et al., 2010), and the proneural protein ASCL1, which is an important regulator of neural fate commitment in neuronal progenitors (Guillemot et al., 1993). The mRNA was induced less strong (3.8-fold ate D11 compared to 8.5-fold induction of circSLC45A4) and additionally increased in expression when neurons were formed (Figure 17 – C). These observations indicate that circSLC45A4, despite its low copy number, may play a functional role in early neuronal differentiation.

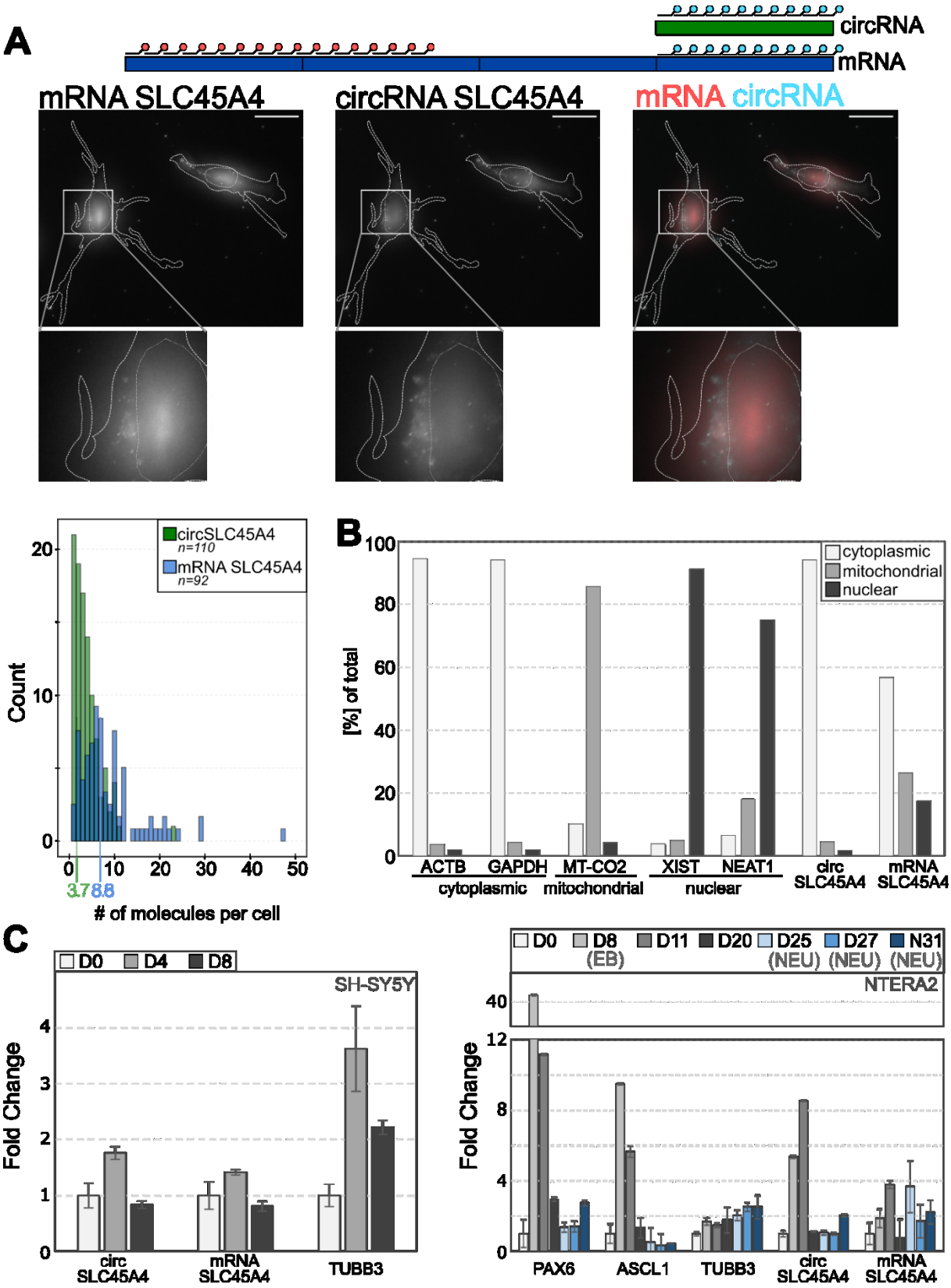


Figure 17: Cytoplasmic circSLC45A4 is dynamically expressed during differentiation of SH-SY5Y and NTERA2 cells **(A)** – Single-molecule FISH of circSLC45A4 and mRNA SLC45A4, probe design, representative pictures and quantification in SH-SY5Y. Scale bar corresponds to 10 µm. Nuclei are marked with dotted lines. **(B)** – Fractionation of SH-SY5Y cells shows that circSLC45A4 is localized to the cytoplasm, while mRNA SLC45A4 is found in the cytoplasm, in mitochondria and in the nucleus. The successful fractionation of SH-SY5Y cells was validated with ACTB and GAPDH as cytoplasmic markers, mitochondrial CO2 (MT-CO2) as mitochondrial marker and NEAT1 as nuclear marker. **(C)** – circSLC45A4 expression during SH-SY5Y and NTERA2 differentiation. Expression of important marker genes that highlight the progression of differentiation are shown. EB – embryoid body stage, NEU – neurons.

7.2.3 Knockdown of circSLC45A4 causes spontaneous neuronal differentiation

The dynamical expression pattern of circSLC45A4 throughout differentiation, indicate a possible functional role of this circRNA within differentiation. Thus, RNAi (Elbashir et al., 2001a; Fire et al., 1998; Meister et al., 2004) was used to specifically downregulate circSLC45A4 in SH-SY5Y cells. In order to specifically target the circRNA and not the linear transcript, the siRNAs were designed to target the circRNA-specific head-to-tail junction (Figure 18 – A). To increase the reliability of the results, three independent but slightly shifted siRNAs were used. Only consistent changes across these three replicate experiment setups were considered. Additionally, to control for effects the transfection of the cells might have, control samples were transfected with a scrambled siRNA that does not have any targets in the known human transcriptome but has a comparable nucleotide composition as the siRNAs targeting circSLC45A4. To monitor the reaction of SH-SY5Y cells to the knockdown of circSLC45A4, the levels of beta-III-tubulin (TUBB3) as important marker for neuronal differentiation were measured by qPCR. Strikingly, TUBB3, proneural factor ASCL1, key regulator of neurogenesis NOTCH1 and mature neuron-characterizing GABBR1 and VGLUT1 were consistently upregulated following the depletion of circSLC45A4 (Figure 18 – B). This indicates that the depletion of circSLC45A4 spontaneously induces neuronal differentiation. While the pre-mRNA levels of SLC45A4 were slightly increased, the levels of mRNA were unaffected after knockdown of circSLC45A4.

This initial result was further supported by immunofluorescence studies of TUBB3 in SH-SY5Y cells after knockdown of circSLC45A4. Again, a consistent upregulation of this marker for the generation of immature postmitotic neurons was observed, while a marker for progenitor cells (GFAP) was constant (Figure 18 – C). One hallmark of neurons is their intricate morphology, exhibiting long protrusions from the cell body, known as neurites or more specifically as axons and dendrites. Hence, after knockdown of circSLC45A4 microscopic pictures were used to trace the present neurites and thereby measure their length. Remarkably, the neurite length was almost twice as long in cells in which circSLC45A4 levels were reduced. The average neurite length was increased to 19.9 µm, compared to 12.0 µm after scrambled knockdown. If SH-SY5Y cells are differentiated by addition of ATRA the neurites grew to an average of 26.7 µm (Figure 18 – E). Thus, downregulation of circSLC45A4 seems to induce neurite growth almost as strong as the standard differentiation protocol.

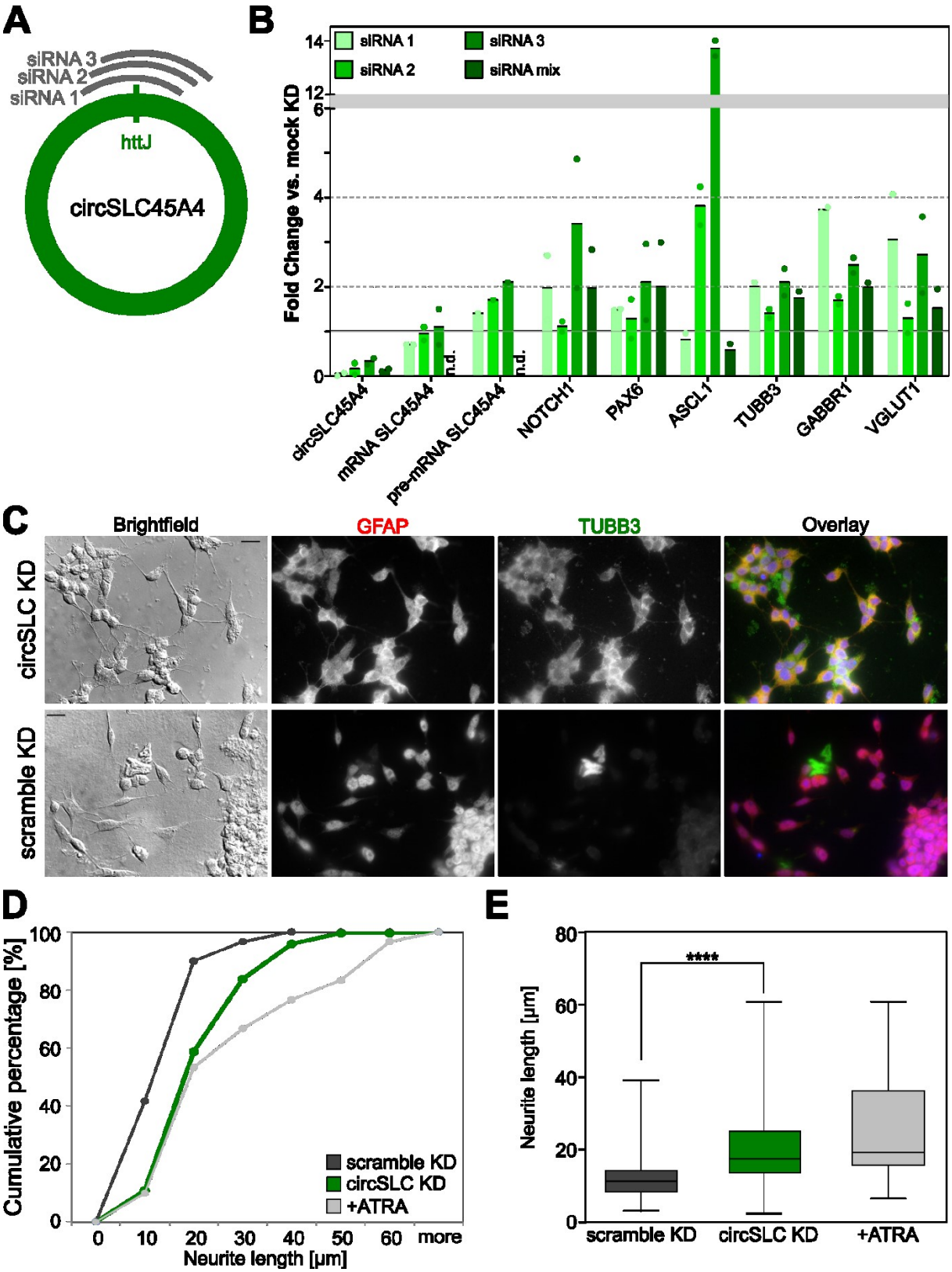


Figure 18: Knockdown of circSLC45A4 in SH-SY5Y cells causes spontaneous differentiation.
(A) – siRNA design strategy for circSLC45A4-specific knockdown. **(B)** – qPCR measurements after knockdown of circSLC45A4 in SH-SY5Y cells, each dot indicating a separate biological replicate, averages are marked. Fold Changes are compared to scramble KD, the baseline is marked as horizontal line at fold change 1. n.d. – not determined **(C)** – Immunofluorescence of SH-SY5Y cells 96 h after knockdown of circSLC45A4, staining for GFAP and TUBB3. **(D+E)** – Results of neurite length measurements after circSLC45A4 knockdown, scramble knockdown and differentiation with ATRA. All neurites of 30 cells per condition measured. **** - t-test p-value < 0.0001

Having established that the knockdown of circSLC45A4 induces spontaneous differentiation of SH-SY5Y cells towards a more neuronal phenotype, RNA-seq was used to understand the transcriptomic changes caused by this perturbation. For this purpose, 6 biological replicates of

circSLC45A4 knockdown were collected, two for each of the three individual siRNAs. For the scramble KD, 2 biological replicates were collected. PolyA selected libraries were prepared for all samples and subsequently sequenced. Sequencing data was demultiplexed using *bcl2fastq,* mapped to the human genome version hg38 using STAR (Dobin et al., 2013) and counted using the R package GenomicAlignments (Lawrence et al., 2013).

Differential expression analysis was performed using DESeq2 (Love et al., 2014b). 1019 significantly upregulated (= 3 %) and 1101 significantly downregulated (= 3.2 %) genes were identified (Figure 19 – A). In addition to the previously detected expression changes of neurogenesis regulators like ASCL1 and NOTCH1, genes connected to axonogenesis were significantly upregulated, e.g. plexins. To validate these findings by qPCR several significantly deregulated genes were selected and measured on independent samples (Figure 19 – B). The strongest upregulated gene was TIMM17B. TIMM17B codes for a protein that is part of the mitochondrial translocase complex TIM23 (Sokol et al., 2014). Furthermore, PRKCA, protein kinase C alpha, was approximately 2-fold induced after knockdown of circSLC45A4. PRKCA is a well-known protein kinase that participates in many signaling pathways, i.e. phosphorylating GAP43 and thereby regulating axonogenesis (Coggins and Zwiers, 1991).

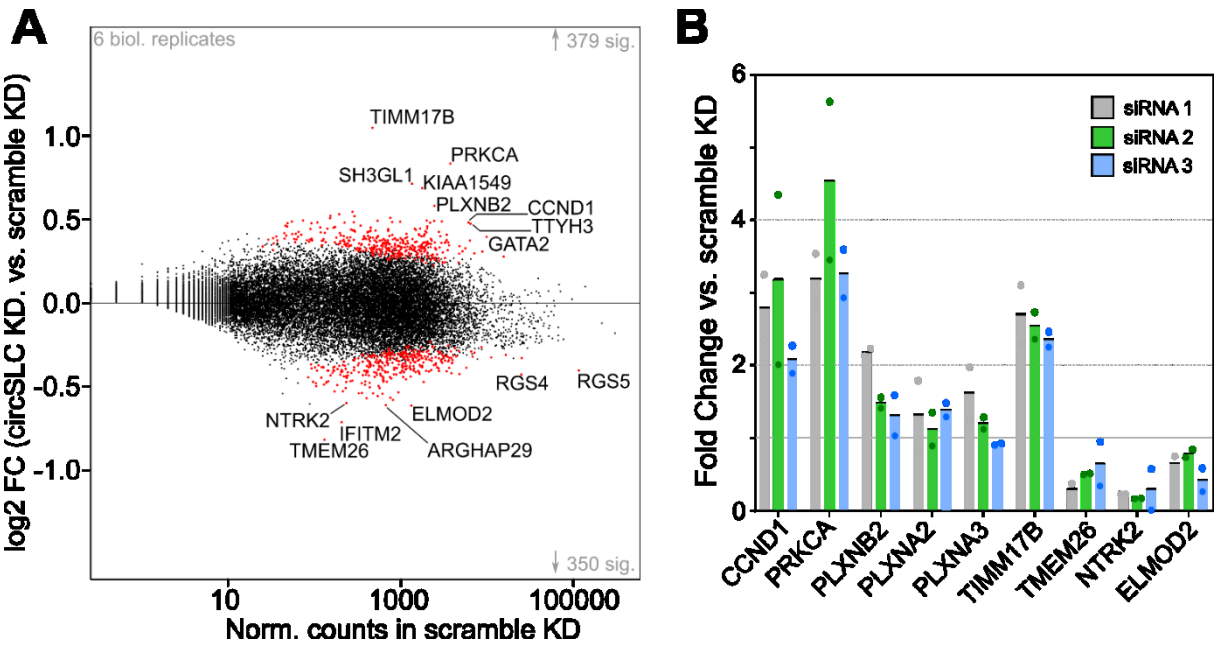


Figure 19: RNA-seq shows that knockdown of circSLC45A4 induces neuron development.
(A) MA plot showing gene expression changes after knockdown of circSLC45A4. Based on 6 biological replicates, two for each siRNA used. Fold changes are considered significant and highlighted in red, if adjusted p-value < 0.05. **(B)** – qPCR validation of selected gene expression changes. Each dot represents one biological replicate. Fold changes are compared to scramble KD, the baseline at Fold Change = 1 is marked as solid horizontal line.

The observation that circSLC45A4 induces spontaneous differentiation was further confirmed by a comprehensive Gene Ontology analysis, using TopGO (Alexa and Rahnenfuhrer, 2016). Genes for GO enrichment analysis were carefully selected, first based on their expression: *i.e.* only genes that had more than 50 counts in the scrambled knockdown were used. Second, only significantly deregulated genes (p-adj < 0.05) were selected and then compared to all genes above the same expression cut-off. GO analysis reveals if amongst the deregulated genes specific GO terms are significantly overrepresented (Fisher's exact test). However, when testing multiple hypothesis, just by chance a significant result may be found. Hence, it is

necessary to adjust the p-value for multiple hypothesis testing, e.g. with the Benjamini-Hochberg method (Benjamini and Hochberg, 1995). This analysis indicated the upregulation of genes connected to homophilic cell-cell adhesion, cell morphogenesis involved in differentiation, neuron projection morphogenesis/guidance as well as axonogenesis (Table 4). Most terms amongst the downregulated genes are evidently not connected to neuronal differentiation, e.g. muscle related terms like smooth muscle contraction (Table 5). However, the GO analysis of downregulated genes is dominated by a few significantly deregulated genes and after Benjamini-Hochberg adjustment of the p-value were no longer significant (p-adj > 0.05).

Table 4: Results of GO term analysis of significantly upregulated genes after circSLC45A4 KD in SH-SY5Y

	Term	Enrichment[1]	p-value[2]	p-adjusted[3]	Example genes
GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	14/65	6.22E-10	1.24E-05	DCHS1, CELSR3, CELSR2, TRO, CADM4
GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules	15/98	2.17E-08	2.05E-04	DCHS1, CELSR3, CELSR2, TRO, CADM4
GO:0007399	nervous system development	72/1573	2.76E-07	1.26E-03	SEMA6C, GATA2, PRKCA
GO:0045652	regulation of megakaryocyte differentiation	9/40	5.14E-07	1.35E-03	SETD1A, KMT2A/B/C, TNRC6B
GO:0050808	synapse organization	18/185	1.03E-06	1.35E-03	SHANK1, L1CAM, NLGN2, BSN
GO:0034968	histone lysine methylation	12/88	2.03E-06	2.35E-03	DOT1L, KMT2A/B/C/D
GO:0016571	histone methylation	13/107	2.88E-06	2.35E-03	DOT1L, KMT2A/B/C/D
GO:0030219	megakaryocyte differentiation	9/51	4.48E-06	2.92E-03	SETD1A, KMT2A/B/C, TNRC6B
GO:0048667	cell morphogenesis involved in neuron differentiation	27/412	5.89E-06	2.98E-03	DAG1, ULK1, SEMA6C, SZT2
GO:0018022	peptidyl-lysine methylation	12/98	6.39E-06	2.98E-03	SETD1A, KMT2A/B/C, TNRC6B
GO:0048812	neuron projection morphogenesis	28/453	1.18E-05	3.01E-03	DAG1, ULK1, SEMA6C, SZT2
GO:0031175	neuron projection development	36/675	1.78E-05	3.25E-03	DAG1, SEMA6C, NCAM1, LRP1
GO:0120039	plasma membrane bounded cell projection morphogenesis	28/465	1.91E-05	3.93E-03	DAG1, SEMA6C, NCAM1, LRP1
GO:0048858	cell projection morphogenesis	28/468	2.14E-05	3.94E-03	DAG1, SEMA6C, NCAM1, NEO1
GO:0048666	neuron development	38/744	2.69E-05	4.28E-03	DAG1, SEMA6C, NCAM1, NEO1
GO:0097485	neuron projection guidance	15/171	2.86E-05	4.54E-03	DAG1, SEMA6C, NCAM1, PLXNA2
GO:0007409	axonogenesis	22/328	3.09E-05	4.54E-03	DAG1, SEMA6C, NCAM1
GO:0045814	negative regulation of gene expression, epigenetic	10/80	3.16E-05	5.28E-03	PHF1, DOT1L, HIST2H2AA3/4
GO:0061564	axon development	23/353	3.22E-05	5.43E-03	DAG1, SEMA6C, NCAM1, PLXNA2

[1] enrichment = number of deregulated genes/number of all genes in that term; [2] p-value determined by Fisher's exact test; [3] adjusted with Benjamini-Hochberg method

Table 5: Results of GO term analysis of significantly downregulated genes after circSLC45A4 KD in SH-SY5Y

	Term	Enrichment[1]	p-value[2]	p-adjusted[3]	Example genes
GO:0060087	relaxation of vascular smooth muscle	4/6	5.97E-06	0.078	GUCY1A3, PRKG1, ADORA1, RGS2
GO:0045932	negative regulation of muscle contraction	5/14	1.73E-05	0.109	GUCY1A3, PRKG1, ADORA1, PDE5A
GO:0044557	relaxation of smooth muscle	4/8	2.68E-05	0.109	GUCY1A3, PRKG1, ADORA1, RGS2

GO:0045986	negative regulation of smooth muscle contraction	4/9	4.72E-05	0.161	GUCY1A3, PRKG1, ADORA1, RGS2
GO:0006041	glucosamine metabolic process	3/4	6.44E-05	0.188	GNPDA2, GNPNAT1, PGM3

[1] enrichment = number of deregulated genes/number of all genes in that term; [2] Fisher's exact test; [3] adjusted with Benjamini-Hochberg method

7.2.4 Pulldown of circSLC45A4 identifies direct interaction partners

In order to understand the role of circSLC45A4 in more detail, its direct interaction partners need to be identified. This can be achieved by crosslinking the RNA to proteins or RNAs in close proximity by UV light (254 nm) and subsequent pulldown (Engreitz et al., 2015). The pulldown is achieved by using multiple biotinylated DNA probes that are antisense to the target. These biotinylated probes can be efficiently enriched with a streptavidin matrix. Interacting proteins can be identified by mass spectrometry analysis, interacting RNAs by next-generation RNA sequencing.

To achieve enrichment of circular RNAs with antisense probes is not trivial. As shown above more than 50 % of human and mouse circRNAs overlap with 5'UTR, 3'UTR or coding sequence of annotated genes (cf. Figure 9 – E). Therefore, probes designed antisense to a circRNA will also enrich the corresponding linear transcript. To make sure that the linear transcript was specifically depleted, samples were treated with RNase R, an efficient exonuclease that leaves circular RNAs intact but degrades linear RNAs (Figure 20 – A).

$100x10^6$ SH-SY5Y cells were UV-crosslinked at 254 nm, lysed, RNase R treated and incubated with biotinylated antisense probes to enrich for circSLC45A4 and its interaction partners. First, the enrichment of circSLC45A4 was determined by qPCR and compared to either an unspecific RNA, actin beta (ACTB), or the linear transcript mRNA SLC45A4. While almost all circular RNA SLC45A4 present in the input was recovered in the eluate, only 1 % of actin beta and 84 % of mRNA SLC45A4 were precipitated. In the negative control, using probes against GFP, which is not expressed in these samples, around 1 % of the initial amount of all three measured RNAs was left in the eluate. Furthermore, the expression level of circSLC45A4 is far lower in the input than that of actin beta and mRNA SLC45A4. Therefore, the relative abundance of circSLC45A4, actin beta and mRNA SLC45A4 was calculated compared to a spike-in. This revealed that after pulldown of circSLC45A4 was 100 times more abundant in the eluate than actin beta or mRNA SLC45A4. In the control pulldown the relative abundances of the measured RNAs to each other were similar to the input (Figure 20 – B). In total, 6 biological replicates were analyzed by mass spectrometry, identifying 961 enriched proteins and potential interaction partners (Figure 20 – C).

Several interesting putative interaction partners were then validated using reverse pulldown experiments. SKP1, an essential component of the SCF complex that is regulating the degradation of cyclin E and NOTCH1 (Nakayama and Nakayama, 2006), was overexpressed in HEK293 cells with a 3xFLAG-tag. The pulldown of SKP1-3xFLAG from HEK293 cells worked efficiently as SKP1 was enriched in the eluate, while the negative control GAPDH was not detected in the eluate (Figure 20 – D). However, the enrichment of circSLC45A4 was marginal and did not change between the samples with and without crosslink at 254 nm. Compared to the IgG negative

control the enrichment is only two-fold and hence the interaction between circSLC45A4 and SKP1 does not seem to be very prevalent. As only 1 % of all circSLC45A4 molecules present in HEK293 lysate were bound to SKP1-3xFLAG. In SH-SY5Y approximately 4 copies of circSLC45A4 are expressed per cell (cf. Figure 17 – A) and expression levels determined by qPCR suggests a similar expression level of circSLC45A4 in HEK293 cells (data not shown). Hence, an interaction of SKP1-3xFLAG and circSLC45A4 would only occur in 1 out of 100 cells. For the other candidate interaction partners, antibodies suitable for immunoprecipitation of the endogenous proteins were available. Therefore, endogenous reverse pulldown experiments for DOCK7, FXR2 and SARM1 were conducted with crosslink at 254 nm from SH-SY5Y cell lysate and monitored by Western blotting (Figure 20 – E, F and G). While the reverse pulldown for DOCK7 and FXR2 worked efficiently, SARM1 could not be pulled down, although pulldown conditions were optimized (several input amounts and several antibody amounts tested, Protein A and Protein G beads tested; data not shown). GAPDH, which was only detected in the input samples, as well as unspecific IgG raised in rabbits served as controls for specific binding of the used antibodies.

DOCK7 is a guanidine nucleotide exchange factor that is involved in neuronal polarity and axon generation through interaction with Rac1 GTPase and myosin VI and was shown to regulate neurite formation (Sobczak et al., 2016). Although circSLC45A4 is 6-fold enriched in the DOCK7 pulldown compared to the IgG negative control, ACTB and mRNA SLC45A4 are 120-fold and 33-fold enriched respectively. Thus, while circSLC45A4 is bound by DOCK7, this interaction is neither exclusive nor prevalent. Given that only an average of 4 molecules (cf. Figure 17 – A) of circSLC45A4 are expressed in SH-SY5Y cells, pulling down only 2.3 % (Figure 20 – E) of circSLC45A4 that existed in the input sample would mean that this interaction exists in only 1 out of 44 cells. Hence, despite interaction of circSLC45A4 with DOCK7, it is unlikely that circSLC45A4 is functioning through the interaction with DOCK7.

FXR2 is an RNA-binding protein that was shown to regulate adult hippocampal neurogenesis, for example through destabilizing Noggin mRNA and hence decreasing BMP signaling (Guo et al., 2011), but also by other non-discovered mechanisms. circSLC45A4 is 14-fold enriched in the FXR2 pulldown, compared to the IgG negative control. But still, only 5.3 % (Figure 20 – F) of circSLC45A4 present in the input sample were pulled down with FXR2. This would translate to one complex of circSLC45A4-FXR2 per 19 cells, given the low expression level of circSLC45A4. In addition, other RNA molecules, mRNA ACTB and mRNA SLC45A4, are again enriched stronger– 155-fold and 116-fold respectively. Thus, an interaction between FXR2 and circSLC45A4 is neither exclusive nor abundant. This renders it unlikely that circSLC45A4 functions through the interaction with FXR2, despite the fact that the interaction was validated by reverse pulldown.

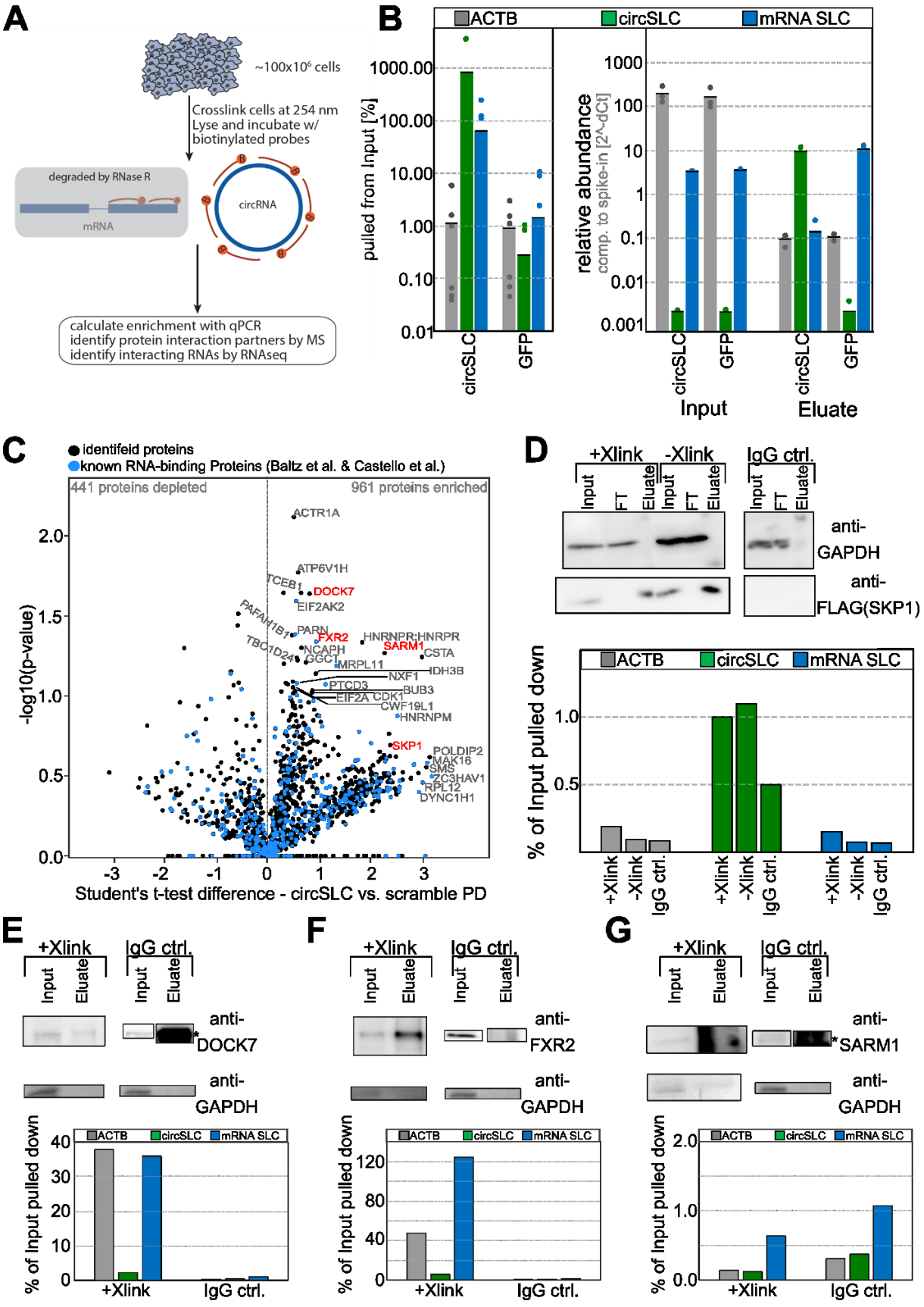


Figure 20: Pulldown of circSLC45A4 identifies direct interaction partners.

(A) – Workflow for a circRNA pulldown experiment. **(B)** – Pulldown efficiency for circSLC45A4, mRNA SLC45A4 and unrelated mRNA ACTB. 6 biological replicates for specific circSLC45A4 pulldown probes and unspecific control probes against GFP. **(C)** – Many potential protein interaction partners were identified by mass spectrometric analysis after circSLC45A4 pulldown compared to the control pulldown, 961 enriched proteins. **(D)** – Reverse pulldown of SKP1 +/- Xlink and compared to IgG neg. control. **(E)** – Reverse pulldown of DOCK7, FXR2 **(F)** and SAMR1 **(G)** + Xlink and compared to IgG neg. control. * - contamination with non-reduced/reduced antibody chains overlays signal.

7.2.5 CircSLC45A4 interaction with miRNAs

Many examples exist in which circular RNAs regulate miRNAs levels and thereby regulate cellular processes. Best characterized is the interaction of Cdr1as with miR-7 (Hansen et al., 2013; Memczak et al., 2013; Piwecka et al., 2017): Knockout of Cdr1as caused downregulation of miR-7 in mouse brain and thereby probably derepressed miR-7 targets, e.g. the immediate early genes c-Fos (Zhao et al., 2015) and Klf4 (Okuda et al., 2013).

Hence, a method that is able to unambiguously identify miRNA:RNA target interactions (Grosswendt et al., 2014) was used to identify whether circSlc45a4 interacts with miRNAs from an AGO HITS-CLIP experiment in mouse brain (Moore et al., 2015). This analysis predicted several miRNAs to interact with circSlc45a4: miR-9-5p, miR-30c/e-5p, miR-135a/b-5p and miR-181a-5p. Strikingly, the interactions were not distributed along the entire circRNA but were predicted to occur in almost the same site on the circRNA (Figure 21 – A). Since comparable human brain datasets are not available, miRNAs that possibly bind to human circSLC45A4 were predicted *in silico* using miRanda (Enright et al., 2003). Again, many miRNA interaction partners of human circSLC45A4 were predicted. However, of the previously identified chimeric miRNA:circRNA interactions only miR-135a-5p was predicted as well (Figure 21 – B). Subsequently, miRNA levels were measured using qPCR and small RNA-seq after knockdown of circSLC45A4 in SH-SY5Y cells (Figure 21 – C and D). Surprisingly, of the predicted interacting miRNAs, only miR-135 was downregulated after knockdown of circSLC45A4.

Small RNA sequencing was used additionally to monitor the levels of both miRNA isoforms, 3p (arises from the 3'-arm of the hairpin) and 5p (arises from the 5'-arm of the hairpin). For miR-135 only the levels of the 5p strand were lower, while the 3p strand was stably expressed after circSLC45A4 KD (Figure 21 - D). This suggests that circSLC45A4 either regulates miR-135 levels post-transcriptionally or that miR-135 levels are changed as a secondary effect due to the spontaneous differentiation that is induced in SH-SY5Y cells after knockdown of circSLC45A4. Hence, we looked at potential tailing and trimming of the miRNA as this would be a possible mode of action for circSLC45A4 to regulate miR-135 levels. For tailing and trimming to occur the miRNA and the target RNA have to be highly complementarity beyond a common 8-mer seed interaction (Ameres et al., 2010; Rüegger and Großhans, 2012). Hence, we used RNAhybrid to predict the interaction structure of miR-135a/b with circSLC45A4 (Krüger and Rehmsmeier, 2006). The pairing between miR-135 and circSLC45A4 (or miR-135b/circSlc45a4) is extending beyond the 8-mer seed interaction. Either the pairing is continued until miRNA base 13 (circSLC45A4 and miR-135a, circSlc45a4 and miR-135b) or additional bulges are introduced at miRNA base 9 (circSLC45A4 and miR-135b – 9 bases, circSlc45a4 and miR-135a – 4 bases) (Figure 21 – E). In any case, the pairing continues at the 3'most part, miRNA bases 19-23 are always paired to the circRNA. However, this base pairing pattern is not sufficient to apply a more general rule for the consequences of this pairing. Degradation is described for extensive base pairing or for partial pairing with restrictive rules (de la Mata et al., 2015) that are not fulfilled in the pairing of circSLC45A4 and miR-135. Yet, small RNA sequencing data was used to investigate potential tailing or trimming of miR-135a-5p after circSLC45A4 knockdown. But no extensive tailing or trimming of miR-135a-5p after circSLC45A4 knockdown was found (Figure 21 – F). The above findings would suggest

that circSLC45A4 either stabilizes this miRNA, catalyzes its biogenesis, or that miR-135 levels are deregulated after circSLC45A4 knockdown due to a secondary effect.

To establish the interaction of circSLC45A4 and miR-135 beyond computational predictions and correlations, the previously described circSLC45A4 pulldown approach was used to measure the enrichment of miRNAs. After pulldown of circSLC45A4 miR-135a was slightly but not significantly enriched (Figure 21 – G). Furthermore, no enriched miRNAs were detected, possibly because of the limitation of the method, e.g. low crosslinking frequency that potentially misses many of the 8 nt long miRNA seed interactions.

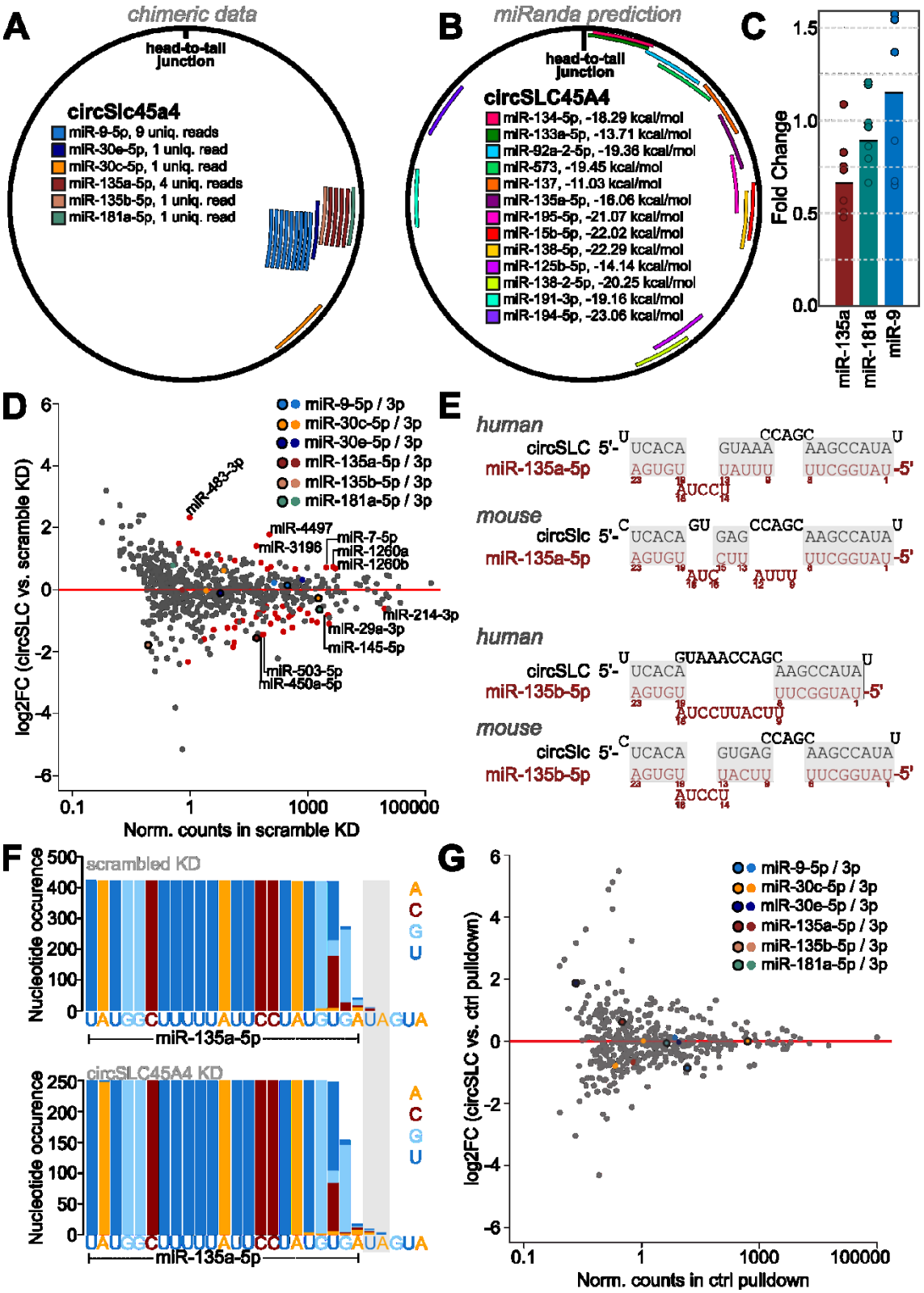


Figure 21: circSLC45A4 interacts with miRNAs.

(A) – miRNA:circSlc45a4 interactions as identified by chimeric reads. **(B)** – miRanda predicts multiple miRNA:circSLC45A4 interactions. **(C)** – Fold changes of miR-135a, miR-181a and miR-9 in SH-SY5Y cells after knockdown of circSLC45A4. Determined by qPCR with TaqMan probes. Average of 7 biological replicates. **(D)** – Analysis of small RNA sequencing confirming the depletion of miR-135a-5p after knockdown of circSLC45A4 in SH-SY5Y cells. Based on 6 biological replicates. **(E)** – Base pairing of circSLC45A4 and miR-135a/b-5p, predicted with RNAhybrid. **(F)** – No extensive tailing or trimming of miR-135a-5p is detected after knockdown of circSLC45A4. **(G)** – Pulldown of circSLC45A4 enriches for certain miRNAs, e.g. miR-135a-5p, but not significantly.

7.2.6 Knockdown of circSlc45a4 *in vivo* significantly disturbs neurogenesis in the developing cortex of mice by depleting the basal progenitor pool

In vitro cell culture models, like SH-SY5Y cells, have the advantage that a complex process like neurogenesis can be analyzed in a simplified system. However, this is also one of their main limitations, as these *in vitro* models lack multiple features that are important in complex tissues or developing organisms, e.g. three-dimensional cell-cell interactions, tissue polarity and complex temporal and spatial regulation of differentiation processes.

To characterize the effect of circSlc45a4 in a more physiologically relevant system, the expression of circSlc45a4 during mouse brain development was investigated (Figure 22 – A). qPCR analyses of samples from different developmental time points of mouse brain development revealed that circSlc45a4 is highest expressed in adult mouse brain. Additionally, it was revealed that circSlc45a4 expression peaks during development at embryonic day (E) 16.5 and at postnatal day 3. The first peak of circSlc45a4 expression is coinciding with peaks of neurogenesis in the nucleus accumbens, tufted cells, isle of Calleja as well as the formation of the corpus callosum and the neurogenesis in cortical layer IV (Finlay and Darlington, 1995). To further assess the distribution of circSlc45a4 within the tissue, E15.5 mouse cortices were stained for circSlc45a4 and mRNA Slc45a4 using BaseScope, an *in situ* hybridization method that allows the specific detection of splice sites (Figure 22 – B). Interestingly, the expression of circSlc45a4 and mRNA Slc45a4 was not restricted to a specific layer of the cortex which suggests a rather general role of these molecules in the developing cortex.

This observation and the upregulation of circSlc45a4 in differentiating progenitors and neurons compared to proliferating progenitors at E14.5 (M. Dori, F. Calegari personal communication, and (Aprea et al., 2013)) suggests a functional role of circSlc45a4 throughout mouse cortical development. Thus, the levels of both circular and mRNA Slc45a4 in the developing neocortex were perturbed by *in utero* electroporation. *In utero* electroporations can be used to introduce DNA constructs or siRNAs to brain regions adjacent to the liquid-filled ventricle at several points during mouse brain development (Fukuchi-Shimogori and Grove, 2001; Saito and Nakatsuji, 2001; Tabata and Nakajima, 2001). Pregnant mice were anesthetized, and embryos were made accessible so that nanoliters of liquid containing the constructs could be injected. The constructs were then electroporated into the cells of the ventricular zone (VZ) of the developing lateral cortex. *In utero* electroporation experiments were conducted at E13.5 to target both proliferating and differentiating progenitors. shRNA constructs were used to knockdown circSlc45a4 (3 constructs as mix), mRNA Slc45a4 (2 constructs as mix) and as a control a scrambled shRNA was used. All used constructs also included a nuclear GFP marker for microscopic identification of manipulated cells by green fluorescence. Mice were electroporated at E13.5, sacrificed and analyzed at E15.5. First, GFP positive cells were collected by FACS and the efficiency of the knockdowns was assessed by qPCR (Figure 22 – D). circSlc45a4 was efficiently depleted by specific shRNA-mediated KD to approximately 20 % compared to scramble KD, the mRNA level was reduced to 75 %. mRNA Slc45a4 was only depleted to 50 % compared to scramble KD with its specific shRNA and circRNA levels were unaffected.

Next, the distribution of manipulated cells within the different cortical layers was quantified (Figure 22 – E and F). Therefore, coronal sections from the manipulated brains were produced, the existing cortical layers at that developmental time point were identified and the number of manipulated cells in each layer was counted. This number was then normalized to the total number of cells in the quantified area and the distribution across the different layers was calculated. Both, the knockdown of mRNA Slc45a4 and circSlc45a4 disturbed the cortical development. Quantification of GFP+ cells (= electroporated) revealed a significant depletion of cells from the cortical plate (CP), for both circRNA and mRNA KD (Figure 22 - F). Intriguingly, only after knockdown of the circular RNA a significant depletion of cells from the subventricular zone (SVZ) was observed (Figure 22 - F). This raised the question, whether the distorted distribution could be caused by a distortion of progenitor pools. Therefore, the proportions of apical progenitors (AP), basal progenitors (BP), and neurons (NEU), defined by cellular localization in the cortex and expression of Tbr2, the main marker of BP (Englund et al., 2005), were quantified. Apical and basal progenitors were distinguished by basal progenitor-specific staining with Tbr2 (data not shown) and subsequently quantified. This analysis revealed a significant depletion of BPs (Tbr2+ in VZ and SVZ), as well as a significant increase of APs (Tbr2- in VZ) after knockdown of circSlc45a4 compared to scramble knockdown (Figure 22 - G). Apical progenitors are located in the ventricular zone and are undergoing mitosis in this region, while basal progenitors are produced from asymmetric divisions of apical progenitors and are mainly located in the subventricular zone and divide there. Basal progenitors are thought to give rise to neurons only, while apical progenitors can give rise to both neurons and glia (Salomoni and Calegari, 2010). Interestingly, the population of BPs was significantly affected only in the VZ and not in the SVZ (Figure 22 - H), indicating that circSLC45a4 KD specifically impaired the cell-fate switch from AP to BP. This could be caused by a migration defect of the cells, premature differentiation or faster differentiation that limits the time spent as basal progenitor during neurogenesis.

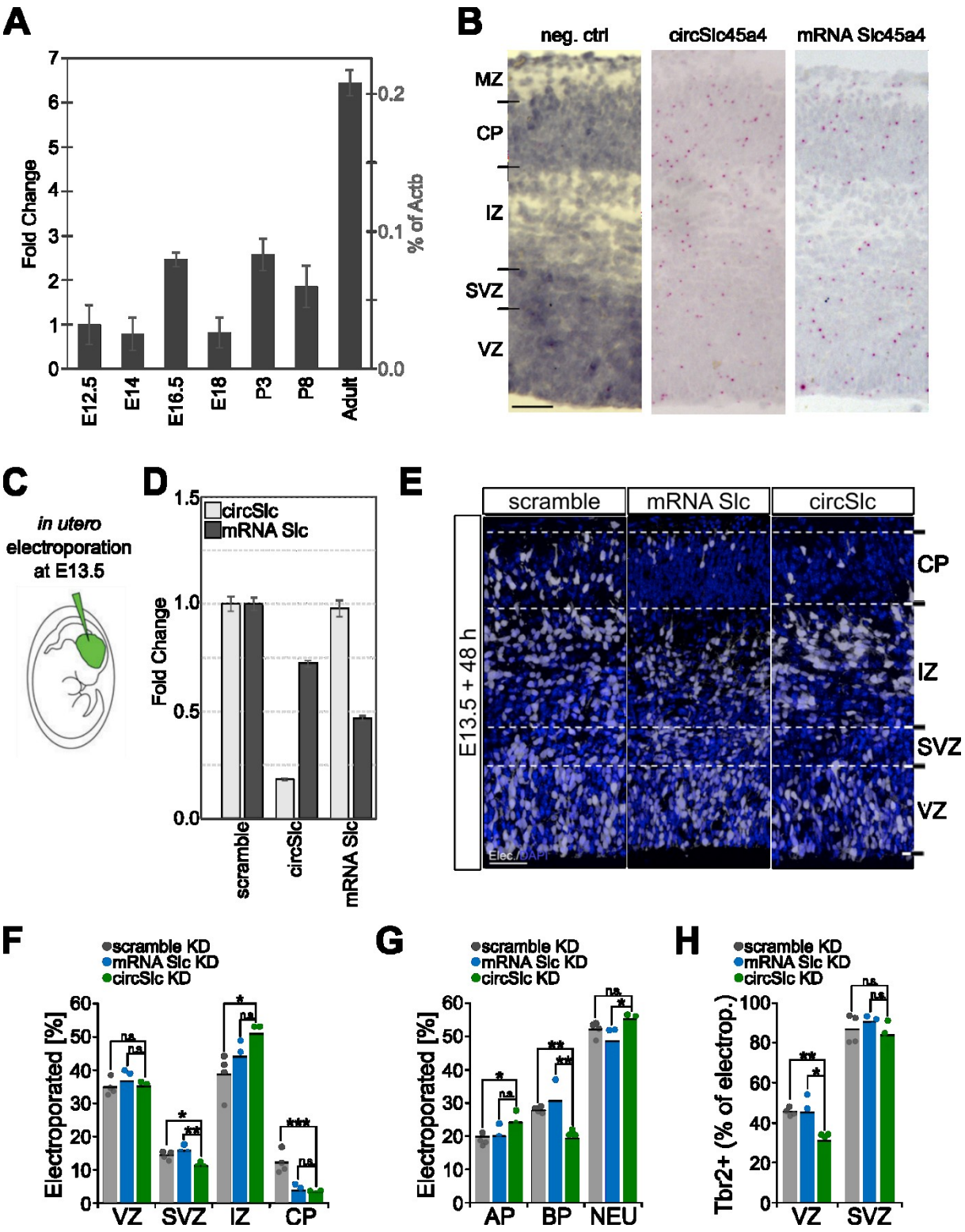


Figure 22: Knockdown of circSlc45a4 disturbs neurogenesis in the developing cortex.
(A) – Expression levels of circSlc45a4 in developing whole mouse brain as determined by qPCR. **(B)** – Splice-site specific *in situ* hybridization of circSlc45a4 and mRNA Slc45a4. Probe against bacterial *DapB* as negative control. MZ – marginal zone, CP – cortical plate, IZ – intermediate zone, SVZ – subventricular zone, VZ – ventricular zone. Scale bar is 50 µm. **(C)** – Representation of an *in utero* electroporation. Subpanel modified from (dal Maschio et al., 2012). **(D)** – Knockdown efficiency of *in utero* electroporation. Determined by qPCR, 3 biological replicates. Error bars as standard errors of the mean. **(E)** – Manipulated cells (white – GFP+) in developing cortex at E15.5 for scrambled control KD, mRNA Slc45a4 and circSlc45a4 knockdown. Nuclei are DAPI stained (blue). Scale bar is 50 µm. **(F)** – Cell distribution per cortical zone, 4 biological replicates. **(G)** – Percentage of electroporated cells that are classified as apical (AP) and basal progenitors (BP) as well as neurons (NEU) by Tbr2+ staining and localization. **(H)** – Percentage of basal progenitors in ventricular (VZ) and subventricular zone (SVZ) after scramble, circRNA or mRNA knockdown.

7.2.7 Knockdown of circSlc45a4 induces significant transcriptomic changes in the developing cortex

The plasmids used for *in vivo* knockdown of circSlc45a4, mRNA Slc45a4 and scrambled control also carried a GFP marker to track the manipulated cells *in vivo*. Furthermore, this enabled the purification of the manipulated cells from embryonic mouse cortices by FACS. RNA was extracted from the cells and used to prepare polyA+ RNA-sequencing libraries. Analysis of the gene expression changes induced after circSlc45a4 knockdown vs. scramble knockdown revealed strong expression changes of up to 16-fold induction or 8-fold downregulation of many genes (Figure 23 – A). shRNAs are prone to induce off-target effects, either by downregulation of unintended targets, unintendedly functioning as miRNAs or simply by overloading Dicer and thereby impairing its function, or by lowering the effective AGO2 concentration available for endogenous miRNAs. To minimize the chance of hitting unintended targets, the shRNA sequences were scored using siSPOTR (Boudreau et al., 2013) which revealed that the potential miRNA seed sequences in the used shRNAs against circSlc45a4 are targeting few (shRNA 1: 4 out of the top-50 downregulated genes, shRNA 2: 1, shRNA 3: 1) of the top-downregulated genes with a maximum of 1 7mer-1A site per target. This renders it very unlikely to induce a downregulation up to 8-fold (Bartel, 2018). Nucleotide BLAST (Altschul et al., 1990) of the 18mer from position 2-19 of the shRNA did not yield any targets with more than 14 nt overlap to the shRNA and hence will not cause a downregulation of these targets either (Elbashir et al., 2001b).

As a scrambled shRNA only serves as control for the overexpression of a short-hairpin RNA in a cell, the knockdown of an expressed target, in this case mRNA Slc45a4, is especially important to ensure that the measured changes are not caused by overloading Dicer or lowering the effective concentration of AGO2. In such an instance, one would expect to find similar changes for both circRNA and mRNA knockdown. To investigate this, principal component analysis (PCA) was applied to the gene expression data that was collected by RNA-seq. PCA linearly transforms the original data while preserving the variation. Principal components (PCs) are directions in the data along which the variation is maximal (Ringnér, 2008). Hence, PCA can be used to visualize the variation in highly dimensional data. Here, PCA showed that circSlc45a4 and mRNA Slc45a4 knockdown in the developing mouse cortex are varying from each other and the scramble knockdown (Figure 23 – B). Furthermore, comparison of circSlc45a4 KD vs. scramble KD with circSlc45a4 KD vs. mRNA Slc45a4 KD revealed that the changes induced by circSlc45a4 knockdown are stable and that these two comparisons are well correlated with R^2=0.6975 (Figure 23 – C). The grey dashed lines mark a log2-fold change between comparisons. Genes that were found to be changed significantly in circSlc45a4 vs. scramble KD (total 3864 genes) are highlighted in red and 96.8 % (= 3740 genes) of those are in the first and third quadrant and below the grey dashed lines. This is a good indication that the measured transcriptomic changes are induced specifically by depleting circSlc45a4. As the up- or downregulation of genes in response to the depletion of circSlc45a4 does not seem to depend on the control the sample is compared to.

The 3864 significantly deregulated genes included interesting candidates involved in neurogenesis. Unc5b, a netrin receptor, is the strongest upregulated gene after knockdown of circSlc45a4, with a 16-fold induction. It is important for the migration of cortical interneurons

from the medial ganglionic eminence (MGE) into the cortex. Unc5b itself is regulated by the transcription factor Sip1, deletion of Sip1 leads to overexpression of Unc5b which in turn strongly inhibits the migration of interneurons from the MGE into the cortex (van den Berghe et al., 2013). Cortical interneurons are a very diverse group of neurons, with at least 23 known subtypes in adult primary visual cortex (Poulin et al., 2016) whose subtypes are already emerging in the developing mouse brain (Mi et al., 2018). One of those subtypes are VIP Parm1+ interneurons. Hence, the observed upregulation of Parm1 after knockdown of circSlc45a4 is especially intriguing. Cortical interneurons originate from the ventral telencephalon and migrate into the cortex (Molyneaux et al., 2007; Wonders and Anderson, 2006), yet the here manipulated cells are derived from the neocortical germinal zone.

It has previously been shown, that the correct migration and positioning of interneurons in the cortex also depends on the interaction with radial glia cells in the neocortical germinal zone (Polleux et al., 2002; Yokota et al., 2007). Genes that are known to regulate the development of the cortex through the production of neurons from progenitor cells in the ventricular and subventricular zones were also deregulated. Most prominently the transcription factor Foxp2, which impairs radial migration of neurons if overexpressed (Clovis et al., 2012). Foxp2 was induced 5-fold after circSlc45a4 knockdown. This upregulation of Foxp2 might explain why the cortical plate is not populated. However, overexpression of Foxp2 does not explain the observed decrease in basal progenitors, as it has been shown that overexpression of Foxp2 at E13.5 does not cause any distortions in the apical or basal progenitor pools (Clovis et al., 2012). Another important upregulated factor is Notch2 (3.8-fold induction), which together with Notch1-4 and their interactors is part of a highly conserved signaling pathway that has an active role in many developmental and homeostatic processes (Bray, 2016). In particular it was shown for the cortex, that overexpression of a constitutively active form of Notch at E13.5 kept progenitors as proliferating radial glia cells and by this prevented the generation of neurons.

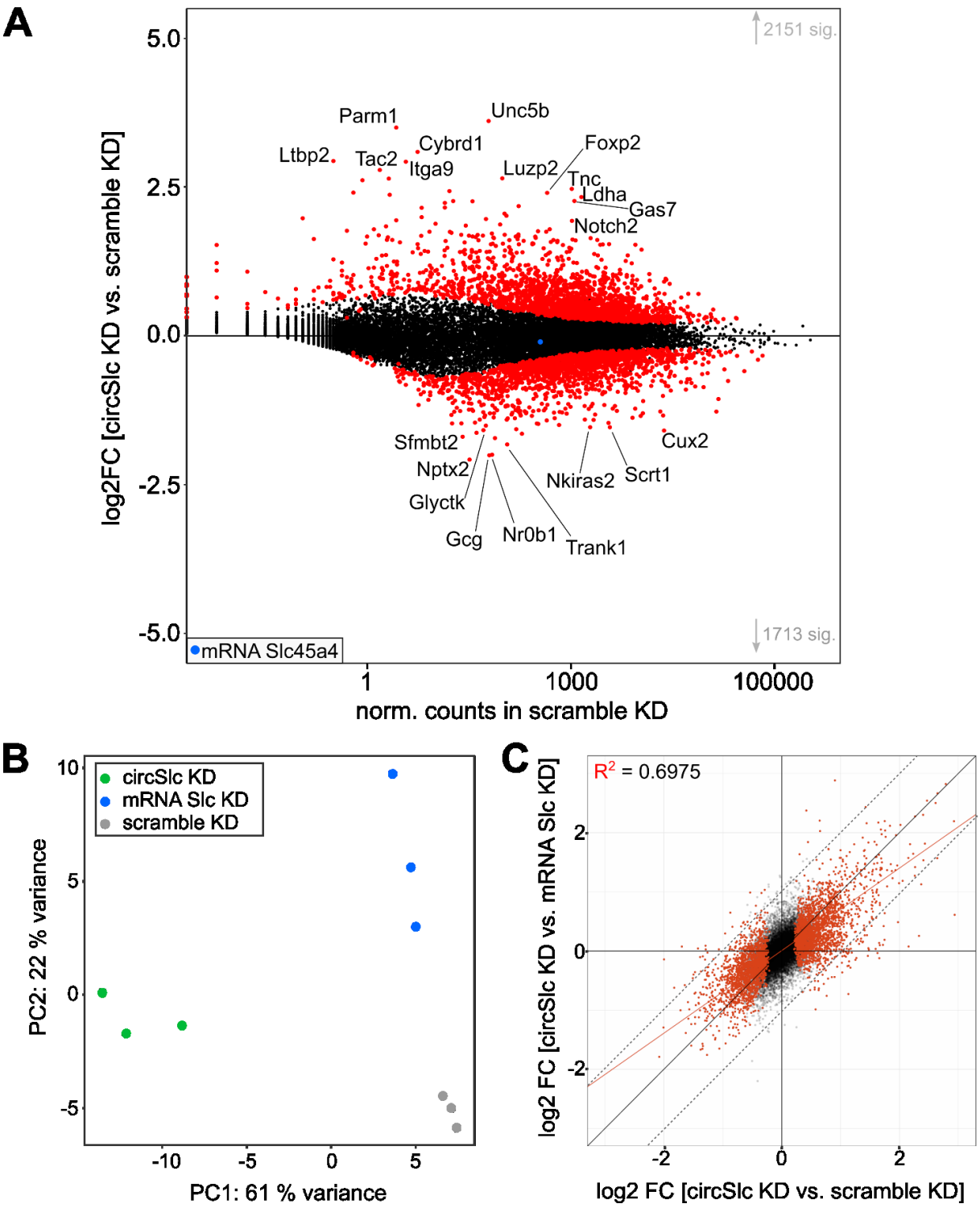


Figure 23: Knockdown of circSlc45a4 and mRNA Slc45a4 are distinct from each other.
(**A**) – MA plot of circSlc45a4 knockdown in mouse cortex versus scrambled knockdown. 2151 genes are significantly ($p < 0.01$) upregulated and 1713 are significantly downregulated, both highlighted in red. (**B**) – Principal component analysis (PCA) separates the three measured conditions. (**C**) – Comparison of circSlc45a4 knockdown vs. scramble KD with circSlc45a4 KD vs. mRNA Slc45a4 KD shows a good correlation (R^2=0.6975, Pearson, red line). Genes that were found significantly changed in circSlc45a4 vs. scramble KD (comp. A) are highlighted in red. Dashed grey lines mark area where log2FC in both comparisons are below 1.

The most strongly downregulated gene is Nptx2 (-4.23 fold), a neuronal pentraxin that is involved in excitatory synapse formation (O'Brien et al., 1999; Tsui et al., 1996; Xu et al., 2003). However, since the cortical plate is not populated after circSlc45a4 knockdown, this is probably a secondary effect. Furthermore, Gcg, the preproprotein of glucagon, is strongly downregulated (-4.0-fold). However, the role of glucagon in the brain, besides its effects on

feeding behavior (Abraham and Lam, 2016), is not understood. Scrt1, a member of the Snail superfamily of transcription factors, is required for the detachment of cells from the ventricular surface and the onset of radial migration (Itoh et al., 2013) and was found to be 2.89-fold downregulated. Reduced levels of this gene could explain the increased number of progenitors in the ventricular zone and the depletion of cells in the cortical plate. Taken together, circSlc45a4 KD induces significant expression changes in developing mouse cortex. However, it remains unclear how circSlc45a4 depletion induces these changes.

Expression changes measured by RNA-seq were further validated by qPCR (Figure 24 - A) and compared to TPM (transcripts per million) values retrieved from RNA-seq experiments (Figure 24 - B). Both of these measures are in good agreement and again suggest that consequences of circRNA and mRNA KD are different and independent from each other.

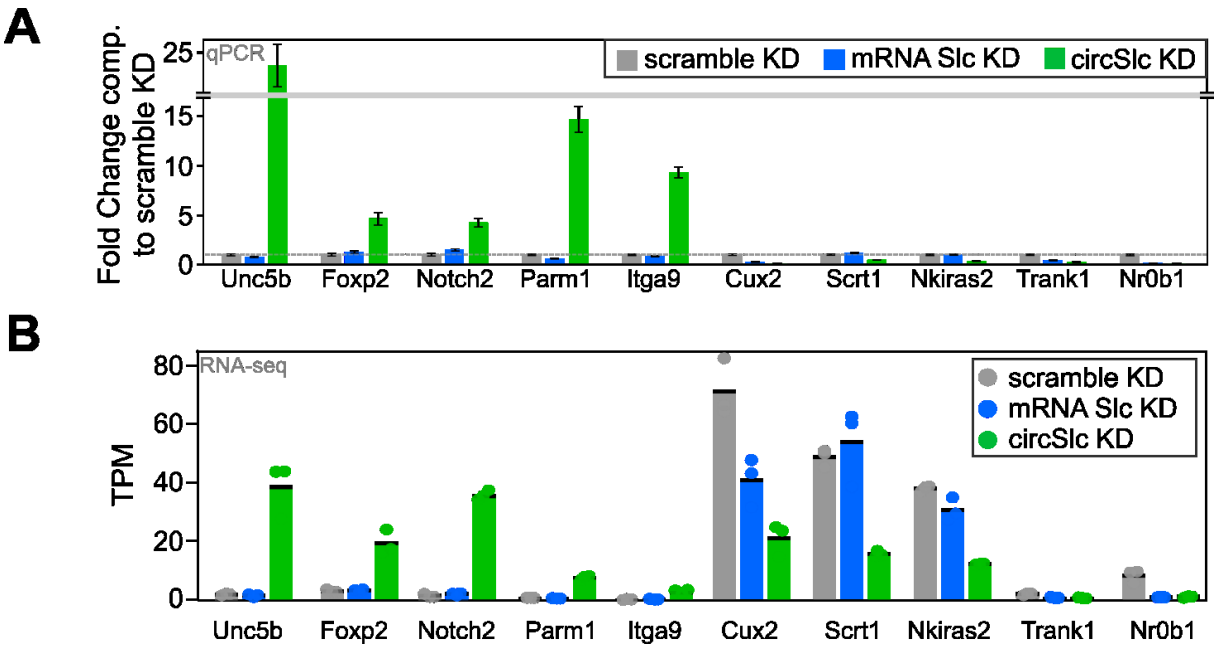


Figure 24: Validation of expression changes in developing cortex with qPCR.
(A) – qPCR validation of induced expression changes after knockdown of circSlc45a4, mRNA Slc45a4 and scramble KD in embryonic mouse cortex. Average of 3 biol. replicates. Standard deviation as error bar. **(B)** – TPM (transcripts per million) values for various genes, illustrating that most changes are specific to circSlc45a4 knockdown.

Instead of looking at deregulated genes individually, it can be useful to search for groups of genes that are associated with certain functions, e.g. with a gene ontology analysis (Table 6). Genes were selected by expression level (at least 50 counts in scramble KD) and significance of the fold change (padj < 0.01). Amongst the upregulated genes those connected to development, cell adhesion and localization as well as signaling are significantly overrepresented. Whilst the downregulated genes are connected to metabolic processes, cell fate commitment and nervous system processes. However, after adjustment of the p-value by the Benjamini-Hochberg method (Benjamini and Hochberg, 1995) none of the GO terms of the downregulated genes remained significant (padj < 0.05), while the GO terms for upregulated genes remained significant.

Table 6: Results of GO term analysis of significantly upregulated genes after circSlc45a4 KD in mouse cortex

Upregulated genes					
	Term	Enrichment[1]	p-value[2]	p-adj[3]	Example genes
GO:0009653	anatomical structure morphogenesis	305/1578	5.00E-19	8.61E-14	Unc5b, Notch2, Pag1, Gnaq, Rbpj
GO:0023052	signaling	513/3060	5.11E-19	8.61E-14	Unc5b, Notch2, Pag1, Gnaq, Rbpj
GO:0007165	signal transduction	466/2761	1.90E-17	3.91E-13	Unc5b, Notch2, Gas7, Foxp2, Fat3
GO:0007154	cell communication	511/3097	2.10E-17	7.12E-13	Unc5b, Notch2, Ldha,Pag1, Plcb1
GO:0023051	regulation of signaling	335/1880	2.41E-15	6.67E-12	Unc5b, Notch2, Plcb1, Gnaq, Igfbp5
GO:0010646	regulation of cell communication	331/1862	5.43E-15	1.38E-11	Unc5b, Notch2, Plcb1, Nell2, Timp2
GO:0048856	anatomical structure development	517/3236	1.01E-14	1.12E-10	Unc5b, Notch2, Gas7, Foxp2, Fat3
GO:0009966	regulation of signal transduction	299/1652	1.68E-14	1.12E-10	Unc5b, Notch2, Plcb1, Sfrp2, Rbpj
GO:0007155	cell adhesion	143/635	3.53E-14	1.99E-10	Fat3, Pag1, Plcb1, Ccdc80, Cyfip2, Vcam1
GO:0035295	tube development	140/617	3.59E-14	1.99E-10	Unc5b, Notch2, Foxp2, Sfrp2, Ptges3, Gna13
GO:0022610	biological adhesion	143/637	4.61E-14	1.12E-09	Fat3, Pag1, Plcb1, Ccdc80, Cyfip2, Vcam1
GO:0032502	developmental process	546/3486	5.57E-14	1.50E-09	Unc5b, Notch2, Gas7, Foxp2, Fat3, Plcb1
GO:0051240	positive regulation of multicellular organismal process	190/937	8.67E-14	1.75E-09	Notch2, Foxp2, Plcb1, Sfrp2, Timp2, Rbpj
GO:0051094	positive regulation of developmental process	174/838	1.27E-13	8.04E-09	Notch2, Foxp2, Plcb1, Sfrp2, Timp2, Rbpj
GO:0035239	tube morphogenesis	117/491	1.60E-13	9.53E-09	Unc5b, Notch2, Sfrp2, Hs2st1, Gna13
GO:0048731	system development	433/2659	3.06E-13	9.53E-09	Unc5b, Notch2, Gas7, Foxp2, Fat3, Plcb1
GO:0032879	regulation of localization	293/1654	5.56E-13	1.07E-08	Plcb1, Gnaq, Igfbp5, Sfrp2, Nell2, Gna13
GO:0051239	regulation of multicellular organismal process	289/1626	5.58E-13	1.67E-08	Notch2, Foxp2, Fat3, Plcb1, Igfbp5, Sfrp2
GO:0007166	cell surface receptor signaling pathway	246/1332	7.36E-13	1.89E-08	Unc5b, Notch2, Usp46, Pag1, Plcb1, Gnaq
Downregulated genes					
GO:0008088	axo-dendritic transport	16/55	9.06E-06	0.094	Tmem108, Nef1, Spg7, Snapin, Ap3m2, Trak1/2
GO:0006221	pyrimidine nucleotide biosynthetic process	9/20	1.69E-05	0.094	Dut, Umps, Cad, Tyms, Dctd, Nme6, Tbpl1
GO:1901607	alpha-amino acid biosynthetic process	14/46	1.85E-05	0.094	Park7, Shmt2, Asns, Atp2b4, Cad, Adi1, Asl
GO:0006220	pyrimidine nucleotide metabolic process	10/25	2.02E-05	0.094	Dut, Umps, Cad, Tyms, Nt5c, Dctd, Nme6, Tbpl1
GO:0008652	cellular amino acid biosynthetic process	14/47	2.43E-05	0.094	Park7, Shmt2, Asns, Atp2b4, Cad, Adi1, Asl
GO:0044281	small molecule metabolic process	131/1105	5.19E-05	0.157	Eno1b, Pgm2l1, Pfkb, Abcg2, Wdtc1, Atp5o
GO:0072527	pyrimidine-containing compound metabolic process	12/39	6.49E-05	0.157	Thtpa, Dut, Umps, Cad, Tyms, Nt5c, Dctd, Nme6
GO:0019752	carboxylic acid metabolic process	72/534	6.53E-05	0.157	Dut, Tyms, Dctd, Dnph1, Shmt1
GO:0072528	pyrimidine-containing compound biosynthetic process	9/24	9.62E-05	0.201	Dut, Umps, Tyms, Dctd, Shmt1
GO:0006082	organic acid metabolic process	74/561	1.06E-04	0.201	Eno1b, Pfkb, Abcg2, Wdtc1, Cbfa2t3
GO:0006520	cellular amino acid metabolic process	32/187	1.15E-04	0.201	Park7, Arg2, Shmt2, Sars2, Asns, Gclc, Azin1
GO:0043436	oxoacid metabolic process	73/557	1.44E-04	0.201	Eno1b, Pfkp, Abcg2, Wdtc1, Cbfa2t3, Park7
GO:0009162	deoxyribonucleoside monophosphate metabolic process	5/8	2.06E-04	0.231	Dut, Tyms, Dctd, Dnph1, Shmt1
GO:0009130	pyrimidine nucleoside monophosphate biosynthetic process	5/8	2.06E-04	0.253	Dut, Umps. Tyms, Dctd, Shmt1
GO:0022616	DNA strand elongation	8/21	2.09E-04	0.253	Parp1/2, Dna2, Rad50, Rnaseh2a, Nbn, Pcna

GO:0009219	pyrimidine deoxyribonucleotide metabolic process	6/12	2.29E-04	0.253	Dut, Tyms, Nt5c, Dctd, Tbpl1, Shmt1
GO:0009157	deoxyribonucleoside monophosphate biosynthetic process	4/5	2.50E-04	0.253	Dut, Tyms, Dctd, Shmt1
GO:0009177	pyrimidine deoxyribonucleoside monophosphate biosynthetic process	4/5	2.50E-04	0.253	Dut, Tyms, Dctd, Shmt1
GO:0036265	RNA (guanine-N7)-methylation	4/5	2.50E-04	0.253	Trmt112, Mettl1, Bud23, Wdr4

[1] number of identified genes in the entire GO term; [2] Fisher's exact test; [3] adjusted with Benjamini-Hochberg method

7.2.8 Single-cell sequencing reveals that circSlc45a4 KD results in depletion of basal progenitors and increase in Cajal-Retzius cells

A few of the observed up- and downregulated genes might be false positives as the initial composition of cell types considered in this analysis was different: *i.e.* in the scramble KD cells from all layers are present, in circSlc45a4 KD the cortical plate and the SVZ are less populated. For example, the downregulation of Cux2 could be explained by the absence of cells from the cortical plate in mRNA and circRNA KD compared to the scramble KD.

Hence, Bseq-SC (Baron et al., 2016), a bioinformatics pipeline that takes single-cell RNA-seq data to inform bulk RNA-seq, was used to assess changes on cell type levels. Bseq-SC can estimate cell type proportions in bulk RNA-seq data based on gene expression sets that have been identified in a corresponding single-cell RNA-seq data set. Hence, the effects on the measured gene expression caused by changes in cell type proportions and by the knockdown of the circRNA or mRNA itself could be deconvolved (Figure 25 - A). Pre-existing single-cell data was taken from (Yuzwa et al., 2017), who assessed different time points of cortical development in embryonic mice, and reanalyzed using Seurat (Butler et al., 2018). To visualize the high-dimensional single-cell RNAseq data, a nonlinear dimensionality reduction method, the t-distributed stochastic neighbor embedding (tSNE) (Amir et al., 2013; van der Maaten and Hinton, 2008), was used. The tSNE plot for embryonic day E15.5 illustrates the 14 identified cell types (Figure 25 - B). Specific marker genes for these 14 cell types were identified and assessed for being exclusively expressed in a single cell type and only marginally in any other cell type (Figure 25 - C). The cluster identity was determined with literature research for described functions of the selected markers. The unique gene expression sets determined for the different cell types present in the cortex at that developmental time point are then used to estimate cell type proportions in the bulk RNA-seq data (Figure 26).

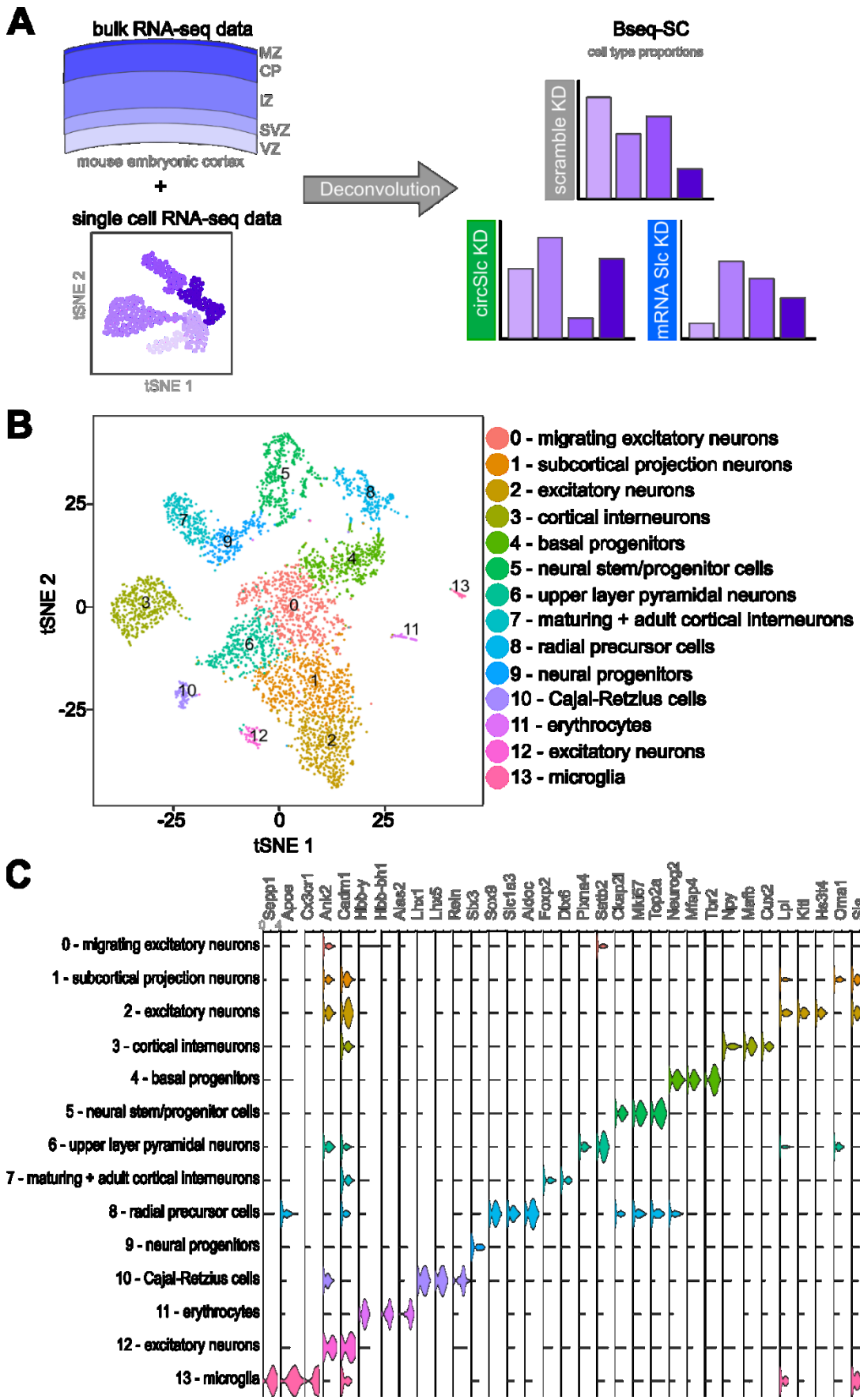


Figure 25: Analysis of single-cell data identifies 14 cell types in mouse embryonic cortex at E15.5.
(A) – general workflow of Bseq-SC. Single-cell RNA-seq data is used to decouple changes introduced by the knockdown from variation in cell populations. **(B)** – tSNE plot of embryonic mouse cortex at E15.5, data from (Yuzwa et al., 2017), reanalyzed. **(C)** – Subset of identified marker genes that were used for cell population analysis. All y-scales are set from 0 to 4.

These marker genes were used to define cell types in Bseq-SC to assess cell type variations after circSlc45a4 knockdown or mRNA Slc45a4 knockdown compared to scramble KD. Consistent with immunofluorescence and cell counting, the basal progenitor population is depleted after knockdown of circSlc45a4, but not after mRNA Slc45a4 knockdown (Figure 26 – A). Furthermore, Cajal-Retzius cell numbers were significantly increased after circSlc45a4 knockdown, but unaltered after mRNA Slc45a4 knockdown. Cajal-Retzius cells were described over 100 years ago by Ramón y Cajal (Cajal, 1890) and Gustaf Retzius (Retzius, 1893) and are a neuronal cell type found in the uppermost layer of the developing neocortex. They produce the extracellular matrix protein reelin (Kirischuk et al., 2014). Reelin is one of the key proteins in regulating radial migration of neurons in the cortex (Gupta et al., 2002). After mRNA Slc45a4 knockdown subcortical projection neurons were more frequent than in the scramble KD control, as well as neural stem/progenitor cells (Figure 26 - B).

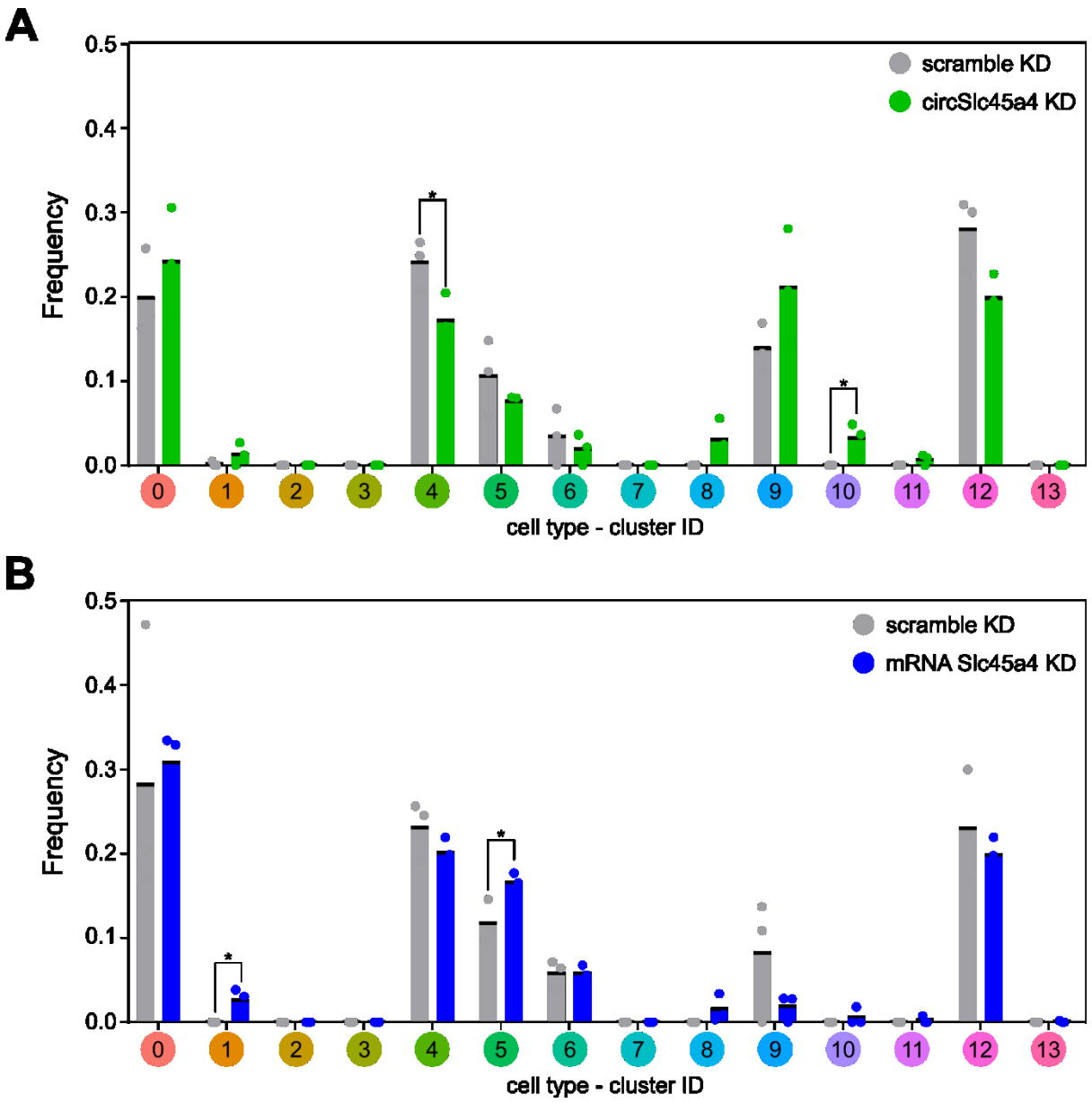


Figure 26: Bseq-SC determines downregulation of basal progenitors and upregulation of Cajal-Retzius cells after circSlc45a4 knockdown.
(A) – Shifts in cell population between scramble KD and circSlc45a4 KD. Cluster 4 (basal progenitors) and Cluster 10 (Cajal-Retzius cells) are significantly altered. t-test: * = $p < 0.05$. **(B)** – Shifts in cell population between scramble KD and mRNA Slc45a4 KD. Cluster 1 (subcortical projection neurons) and cluster 5 (neural stem/progenitor cells) are significantly altered. t-test: * = $p < 0.05$.

The predicted increase of Cajal-Retzius cells is intriguing, as this could cause the depletion of the basal progenitor pool. Hence, reelin immunofluorescence was used to qualitatively assess the number of Cajal-Retzius cells (Figure 27). Reelin levels were strongly increased after knockdown of circSlc45a4 in embryonic mouse cortices at E15.5. Knockdown of mRNA Slc45a4 did not change the reelin levels compared to the scramble knockdown control.

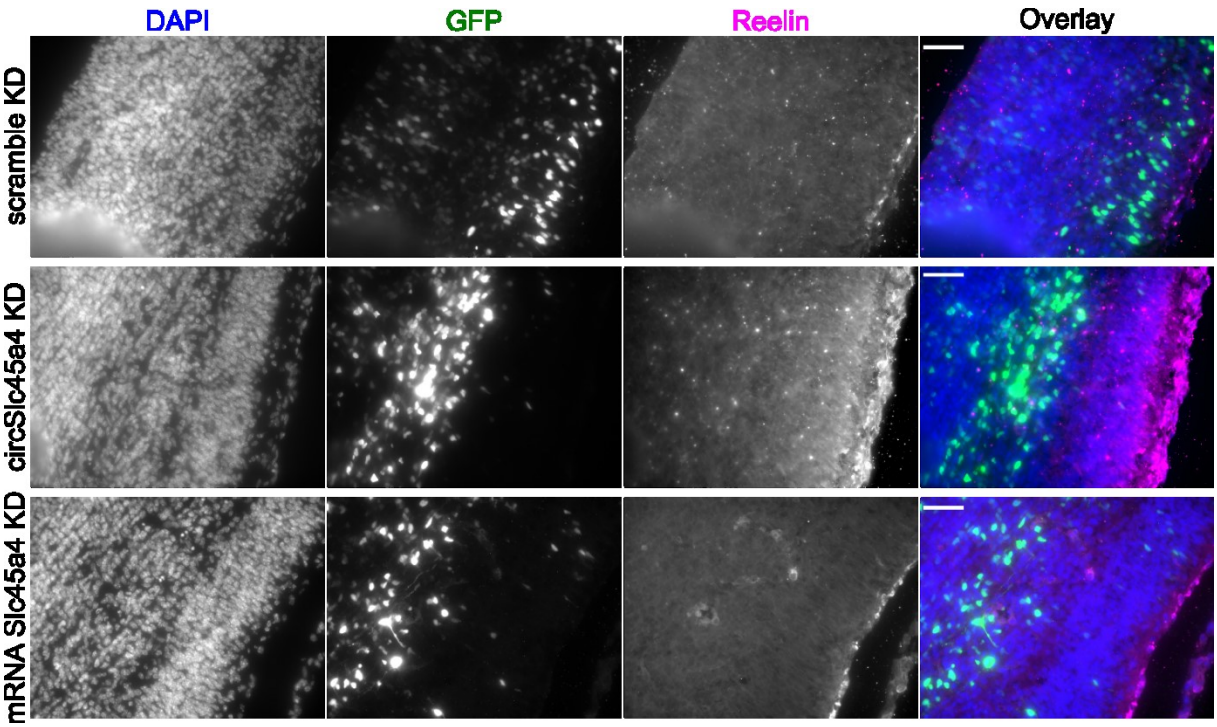


Figure 27: circSlc45a4 knockdown causes an increase in Reelin.
Electroporated mouse cortices at E15.5. Electroporated cells are GFP+, nuclei are stained with DAPI. Reelin levels are increased after knockdown of circSlc45a4 but not after knockdown of mRNA Slc45a4 or in the control knockdown. Scale bar 50 µm.

7.2.9 Investigating the potential translation of circSLC45A4

Although the impact on neuronal differentiation by knockdown of circSLC45A4/circSlc45a4 in different systems, human SH-SY5Y cells and developing mouse cortex, is evident, its mode of action remains elusive. RNA molecules can either act directly on targets, e.g. interactions partners, or they can hold information for protein translation. Importantly, circRNAs have been shown to act as both: directly (e.g. (Memczak et al., 2013; Piwecka et al., 2017) and as translation template (Legnini et al., 2017; Pamudurti et al., 2017).

circSLC45A4 contains the canonical start codon of its corresponding mRNA and a potential open-reading frame (ORF) would extend beyond the head-to-tail junction and terminate with 4 stop codons in the 5'UTR of human SLC45A4 (Figure 20 - A). Next, the conservation of the potential protein sequence after the head to tail junction was analyzed (Figure 20 – B). The protein sequence after the head-to-tail junction shows lower conservation than the part of circSLC45A4 that overlaps with the CDS of the mRNA. However, this part is still exceptionally conserved for a 5'UTR that is not annotated to be translated. Furthermore, the STOP codons found in human are still found in dog and all mammals investigated have at least 2 STOP codons in the very same position. Chicken and frog (*Xenopus tropicalis*) would already

terminate the potential ORF one amino acid after the head-to-tail junction. Still, a translation of circSLC45A4 is conceivable based on its sequence properties.

Therefore, an *in vivo* model was employed to investigate translation in the adult mouse brain. So called Ribotag mice (Sanz et al., 2009) express an HA-tagged version of the ribosomal protein Rpl22. With help of the Cre-LoxP system (Tsien, 2016) this Rpl22-HA is only expressed in certain cell types. In this case all Neurod6-expressing neurons, which are mostly cortical neurons (Figure 20 - C), heterozygously express Rpl22-HA. Brains from these and wildtype (WT) mice are lysed to generate homogenate that is than subjected to an HA-specific pulldown. The retrieved eluate from the WT brains should only contain background binders, while the eluate from the Neurod6-Cre:Ribotag$^{+/-}$ mice should contain RNAs that are bound to ribosomes in Neurod6-positive neurons. To verify the specificity of the pulldown, certain housekeeper, neuronal and non-neuronal marker genes were quantified using qPCR (Figure 20 – C). A fold change compared to the WT eluate larger than 1 would indicate an enrichment of a specific RNA at ribosomes from Neurod6-positive neurons. This was indeed the case for the housekeepers Eef1a and Actb, as well as the neuronal marker genes Neurod6 and Rbfox3 and the ribosomal RNA 18S. As expected, non-neuronal marker genes like Gfap, Aldh1l1 and Olig1 were depleted, or at least not more abundant than in the control pulldown. Having verified the validity of this pulldown experiment, the enrichment of circSlc45a4 and mRNA Slc45a4 was measured. However, none of these two RNAs was enriched at ribosomes from Neurod6-positive neurons (Figure 20 – C).

Yet, this method was developed with non-circular RNAs as main target. If there are circRNA-specific translation mechanisms, like specialized ribosomes, one would not expect to capture a circRNA with the before mentioned Ribotag-pulldown. Hence, the identification of peptides that would be specific for translated circSLC45A4 from SH-SY5Y cell lysates by mass spectrometry was attempted. Therefore, an *in silico* trypsin digest was performed with the ExPASY tool PeptideMass (Wilkins et al., 1997, 1999) and based on this, peptides with at least 5 amino acids length that are specific for the circRNA were selected and ordered as heavy-spike tide synthetic peptides (Figure 20 – B). Since some of the peptides are overlapping with a hydrophobic transmembrane domain also found in the linear mRNA protein product and as further hydrophobic amino acids (V, I, L, F, W, C, L and R) occur after the head to tail junction, SDS extraction was used to prepare lysates from human SH-SY5Y cells and mouse Neuro2a cells. A mass spectrometry setup was used that is able to identify proteins in an amol expression range. However, a circSLC45A4-specific peptide was not found. This could either mean that circSLC45A4 is not translated or that its translation product is lower abundant than in the amol-range.

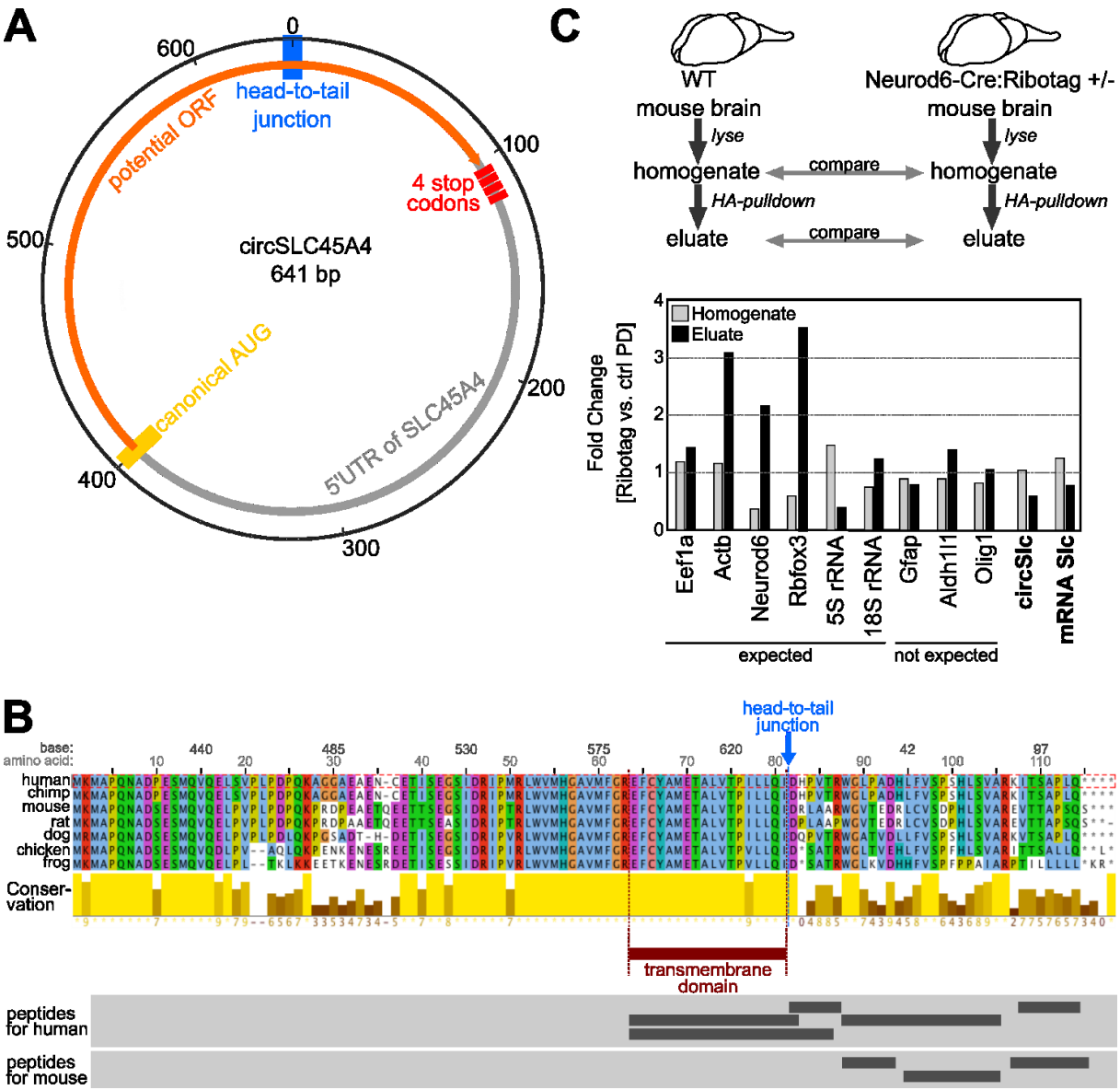


Figure 28: Analyzing the putative translation of circSLC45A4.
(A) – Overview of circSLC45A4, its potential open-reading frame and how it overlaps with the canonical Start codon and part of mRNA SLC45A4. **(B)** – Conservation of the peptide sequence encoded within circSLC45A4 up to frog (*Xenopus leavis)*. Annotated is a transmembrane domain (dark red) that also exists in the protein encoded by the mRNA, the head to tail junction (blue) and the peptides that have been used in mass spectrometry to identify the peptide encoded by the circRNA. **(C)** – qPCR results for a Ribotag pulldown from Neurod6+ cells from adult mouse brain.

7.3 3D-seq: a method to integrate single-cell and spatial transcriptomics

7.3.1 Developing a tool to integrate single-cell and spatial transcriptomics

The already beforehand discussed mammalian cortex is a good example for a tissue whose spatial organization is of utter importance. The cortex is built up in layers originating from waves of migrating neuronal cells. Most current single-cell methods, although delivering a very detailed picture of the transcriptional landscape, do not allow for the recording of the original spatial position of a cell in the tissue. To overcome this problem, two approaches seem feasible: a) leaving the tissue intact and using highly multiplexed single-molecule FISH with pre-defined probe-sets or b) compartmentalizing the tissue and labeling cells *in situ* in any given compartment before breaking the tissue up into single cells. smFISH has a number of limitations, e.g. the used probe set has to be predefined and therefore unbiased investigation of the tissue is impossible, the labeled transcripts can only be assigned to a cell if they are nuclear or at least perinuclear, the turnaround time for each experiment is high (approx.. 14 days for a 5x5 mm piece of mouse cortex with 600 probes, Sten Linnarsson personal communication) and the amount of data collected is large (terabytes per experiment, Sten Linnarsson (Karolinska Institute, Stockholm, Sweden), personal communication) at an overall high cost per experiment (cf. 3.3 - Investigating single-cell gene expression with spatial resolution).

Hence, a tool to preserve spatial information of any given cell while still obtaining transcriptome data with single-cell resolution was developed. For this purpose, a tissue is first compartmentalized to obtain groups of 500-5000 cells depending on the cell density in the tissue (Figure 29 - A): this tissue is fresh frozen in embedding medium as it is usually done for *in situ* hybridization experiments and similar techniques (Fischer et al., 2008). The tissue is then cryosectioned in sections with the thickness of approximately one cell layer, for mouse brain that would correspond to roughly 10 µm. Then, the tissue is fixed and permeabilized using methanol and a 3D-printed grid is applied to the section with the help of a mounting device. The mounting device presses the grid onto the tissue to avoid any leakiness in-between the compartments. At the same time only the tissue under the grid walls gets pressed and subsequently destroyed. Yet, the overall integrity of the tissue section is preserved which would allow several consecutive rounds of *in situ* barcoding to increase the resolution of the method. The *in situ* barcoding can either be achieved by reverse transcribing the RNA and including compartment specific barcodes in the primer sequences or by ligation of short oligonucleotides. After the *in situ* barcoding step a single-cell solution has to be obtained, either enzymatically, or in case this is not feasible, at least the nuclei can be released with a tissue-appropriate method. Next, the cells undergo an optimized combinatorial indexing protocol (Cao et al., 2017; Rosenberg et al., 2018) to obtain transcriptome information with single-cell resolution (Figure 29 - B). The transcriptome information as well as the barcodes attached to the molecules during the protocol are read by next-generation sequencing. This data is then used to create an *in silico* tissue map that traces the cells back to their original tissue position while having information on their transcriptome.

To date, this would be the only existing method that allows the investigation of single-cell transcriptomes while preserving each cell's spatial position in the very same tissue section.

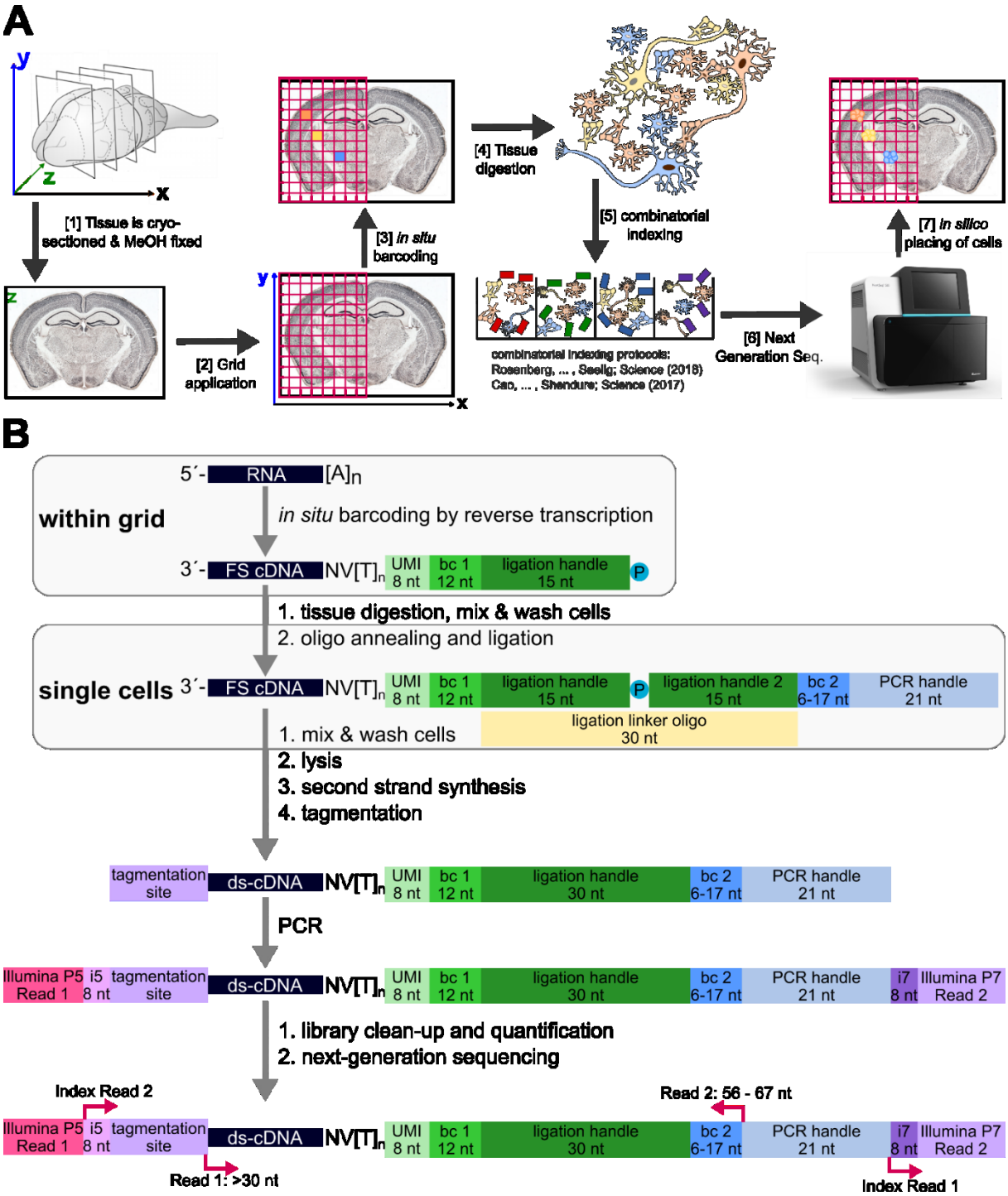


Figure 29: General workflow of 3D-seq.
(A) – **[1]** Tissue (e.g. a mouse brain) is cryosectioned, fixed, and permeabilized in methanol. **[2]** The grid is applied to the tissue, the grid walls section the tissue in compartments. **[3]** Cells in each compartment are barcoded *in situ* with a reverse transcription reaction and an individual barcode per compartment. **[4]** The tissue is enzymatically digested (in case enzymatic digestion is not feasible, nuclei can be released) to obtain a single-cell solution. Each single cell contains cDNA with a compartment specific barcode. **[5]** – Combinatorial indexing is used to obtain single-cell resolution. **[6]** The transcriptome and the attached barcodes are read by next generation sequencing. **[7]** – A tissue map is created *in silico* by placing the cells back in their original tissue position. **(B)** – Molecular workflow for 3D-seq. UMI – unique molecular identifies; bc 1 – barcode 1; bc 2 – barcode 2; i5 and i7 – Illumina multiplex indices.

7.3.2 Establishing *in situ* reverse transcription in fixed tissue for 3D-seq barcoding

The RT reaction is the first step in this multi-step protocol (Figure 29 – B) with the purpose to create a sequencing library that contains information on single-cell transcriptomes and spatial information for each of these cells. Thus, a crucial step of this protocol is the *in situ* barcoding of cells in each grid compartment to record their spatial position within the tissue. Hence, it needs to be determined to which extend an *in situ* barcoding by reverse transcription (RT) is feasible. The barcode would be part of the RT primer (cf. Figure 29 – B "within grid").

Template-switching, also known as SMART (Zhu et al., 2001), has wide applications in single-cell RNA-sequencing where amplification of low input material amounts is especially important. A chimeric RNA/DNA oligo (template switch oligo, TSO) that has 3 Gs on its 5'end can be used to cause a so-called template-switch, where the RT enzyme will continue the replication beginning at the 5'end of the TSO. This way a known sequence is incorporated at the 3'end of the first-strand cDNA that can later be used for additional full-length amplification of the cDNA, e.g. by PCR. The two RT enzymes under investigation are Maxima H- Reverse Transcriptase and Superscript IV. The main difference between these two Moloney murine leukemia virus (MMLV)-based RT enzymes is there template switching capability. Upon reaching the 5'end of an RNA most MMLV RTs will use their terminal transferase activity to add a few nucleotides, mostly C, to the 3'end of the newly synthesized cDNA strand preparing the RT product for template switching. Maxima H- Reverse transcriptase possess this template switching capability (ThermoScientific, 2017), while Superscript IV was engineered to mostly lose this capability (Xu et al., 2017). While the ability to amplify full-length cDNA can be advantageous in any single-cell RNA-seq protocol, processivity and efficiency, understood as ability to reverse transcribe RNAs from a broad expression range, should be considered as well. For this purpose primers were designed to either measure the amount of a reverse transcribed molecule with the entire RT primer attached (Figure 30 – A, Set 1) or to measure the amount of reverse transcribed molecule, e.g. due to self-priming, RT hopping etc. which could cause additional amplification with Set 2 compared to Set 1, (Figure 30 – A, Set 2). These primers were used in a qPCR reaction on cDNA generated *in situ* from different amounts of permeabilized P19 mouse cells (Figure 30 – B). 3D-seq will mostly operate in a range were 100-5000 cells are localized within one compartment. In this range, Superscript IV was significantly more efficient in reverse transcribing the two tested genes Actb (5.3-fold in 1000 cells and 11.1-fold in 100 cells) and Tubb3 (16.6-fold in 1000 cells and 100-fold in 100 cells). Thus, Superscript IV would capture up to 11.1-fold more transcripts from the same number of cells. An advantage that outweighs the option to amplify the material by PCR after template switching with Maxima H-. Hence, Superscript IV was used in all following experiments.

Next, it was analyzed whether permeabilization time and reaction time of the RT influence its efficiency (Figure 30 – C). Therefore, a 10 µm mouse brain slice was permeabilized with methanol and the RNA was reverse transcribed *in situ.* The longer the cells were permeabilized and the longer the RT reaction time was, the higher was the yield for the genes tested in a subsequent qPCR. In total, the yield increased 1.5- to 2.5-fold. Notably, the efficiency was even higher compared to an identical approach, where TRIzol extraction, which is otherwise routinely used for RNA extraction, was applied. A 10 µm mouse brain slice contains little RNA

and TRIzol extraction, especially the precipitation step, is inefficient when used on small RNA amounts. This finding does not allow any conclusions about the efficiency of *in situ* vs. *in vitro* reverse transcription. It rather implies, that more material is recovered after *in situ* RT than by TRIzol extraction from low input samples.

Furthermore, *in situ* RNA is highly structured and bound by proteins. Hence, certain RT pre-treatments and additives were evaluated in a standard reaction volume of 20 µl and a 3D-seq adjusted volume of 5 µl (Figure 30 – D). The effects of protein digestion with a mild Proteinase K treatment, increasing the viscosity of the mixture with Ficoll or resolving the RNA's secondary structures by pre-heating were analyzed. For this purpose, qPCR was used to quantify 4 genes from methanol-permeabilized HEK293 cells: ACTB, TUBB, VCL and TOP1. While these changes to the protocol rather decreased the yield in the larger reaction setup (20 µl), the effects in the smaller reaction volume were either neglectable or limited to 2 out of 4 tested genes. Therefore, none of these adjustments were included in the 3D-seq protocol.

The RT reaction is the first step in a multi-step protocol (Figure 29 – B) with the purpose to create a sequencing library that contains information on single-cell transcriptomes and spatial information for each of these cells. Therefore, it is important to analyze metrics like the number of recovered cells, the median number of genes per cell, the median number of UMIs (unique molecular identifier, (Islam et al., 2014; Kivioja et al., 2011)) per cell, and the PCR bias (= number of reads/UMI) to evaluate this protocol and the impact of its steps by next-generation sequencing (Figure 30 – E). Using Superscript IV and hence not including a template-switch step (HEK293) was compared to using Maxima H- with a template-switch test (HEK293), a standard Drop-seq run with template switch (HEK293) and a SPLIT-seq run with template switch (mouse brain). The median number of genes or UMIs/cell are largely dependent on the number of reads that are spent per cell. So, while the number of genes/UMIs per cell are larger for the Drop-seq and SPLIT-seq runs than for the 3D-seq runs, the numbers become very comparable when considering the number of reads per cell. The PCR bias shows that the 3D-seq protocol variant w/o TSO (= Superscript IV, 10.1 reads/UMI) is comparable to SPLIT-seq, while both protocols need to spend approximately twice the number of reads per cell to obtain a similar number of genes or UMIs/cell as Drop-seq. However, it is evident, that the protocol variant using Superscript IV (10.1 reads/UMI) is significantly more efficient than using Maxima H- (44.9 reads/UMI) in the initial RT step. Taken together, these RT optimization experiments increased the yield of the protocol's first step significantly by 8.8-fold. The choice of the right RT enzyme alone increased the yield 4.4-fold (Figure 30 - E, PCR bias), adjusting the permeabilization and reaction time increased the yield by another 2-fold (Figure 30 – C).

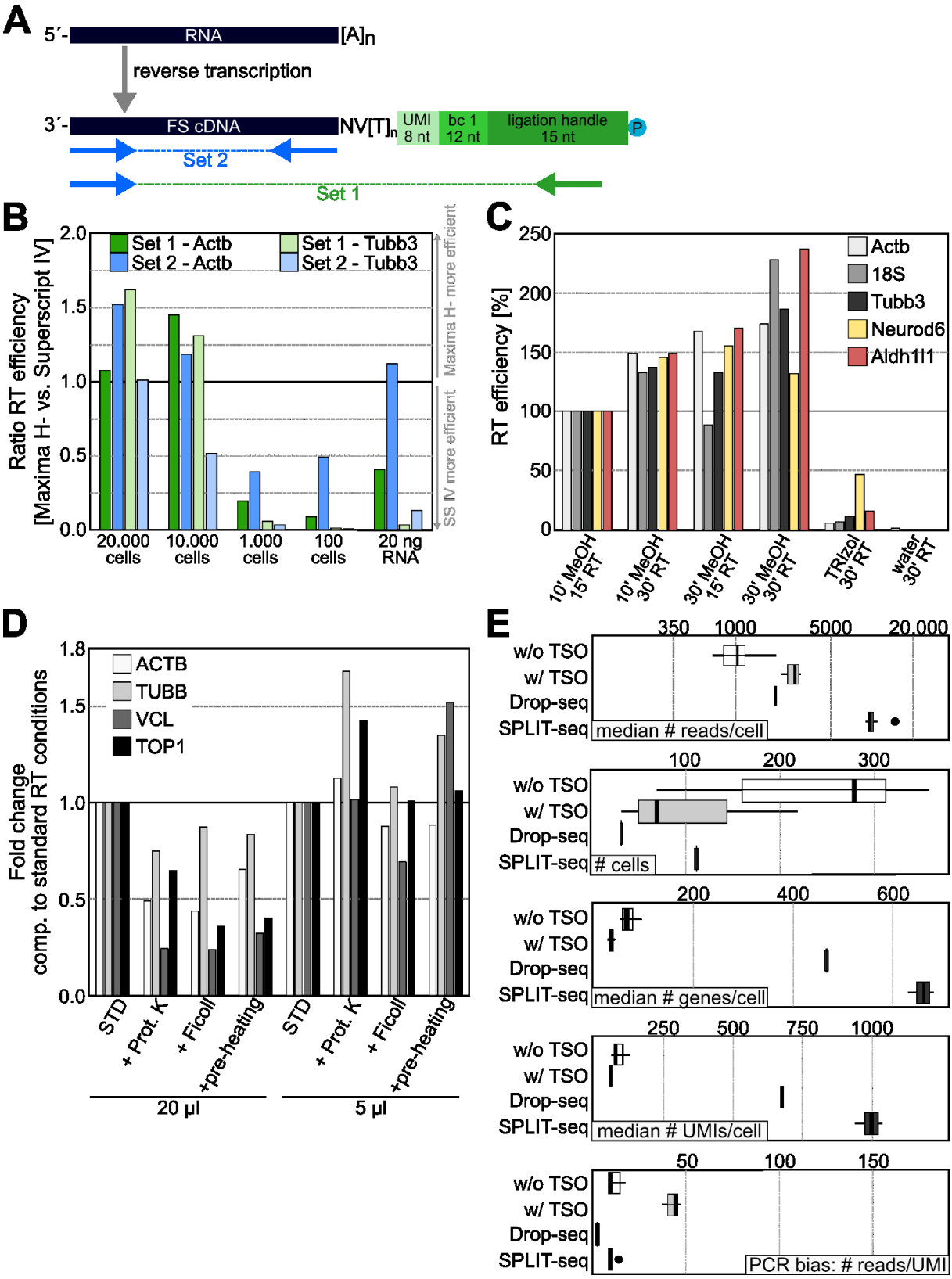


Figure 30: Reverse transcription *in situ* is efficient.

(A) – Primer design scheme for experiment shown in (B). **(B)** – Ratio of RT efficiencies between two different enzymes – Maxima H- and Superscript IV. Values > 1 indicate that Maxima H- works more efficient in that condition, values < 1 indicate that Superscript IV is more efficient. 20 ng of naked RNA (roughly ~1000 cells) were used as proxy for RT efficiency. **(C)** – RT efficiency depending on time of methanol permeabilization (10 or 30 min) and RT reaction time (15 or 30 min) on adult mouse brain slices. One slice was harvested in TRIzol and RNA was extracted for RT, no permeabilization step. Water served as negative control. **(D)** – Evaluation of (pre)-treatment steps for RT. STD – standard conditions without pre-treatment, + Prot. K – mild Proteinase K digest before RT, + Ficoll – Ficoll addition to RT reaction, +pre-heating – heating of permeabilized cells to 55°C, 5 min, before addition of RT enzyme. **(E)** – Evaluation of sequencing results for a 3D-seq run on HEK293 cells. w/o TSO – Superscript IV RT, w/ TSO – Maxima H- RT with addition of a template-switch oligo step after ligation of the second barcode, Drop-

seq – standard Drop-seq run from HEK293 cells, SPLIT-seq – SPLIT-seq run on mouse brain from (Rosenberg et al., 2018). For w/o TSO and w/ TSO only cells with at least 100 UMIs were considered.

7.3.3 Methanol fixed mouse brain can be digested into a single cell solution using papain

Methanol is routinely used to fix and permeabilize cells, for example in Drop-seq experiments (Alles et al., 2017). SPLIT-seq uses PBS/Triton X-100 for permeabilization and para-formaldehyde for fixation. In order to compare the efficiency of both approaches, the recovered median number of genes/cell and UMIs/cell was analyzed. Although the number of reads spent per cell was comparable (not shown), methanol fixation was significantly more efficient in terms of recovering genes and UMIs/cell (Figure 31 – A and B).

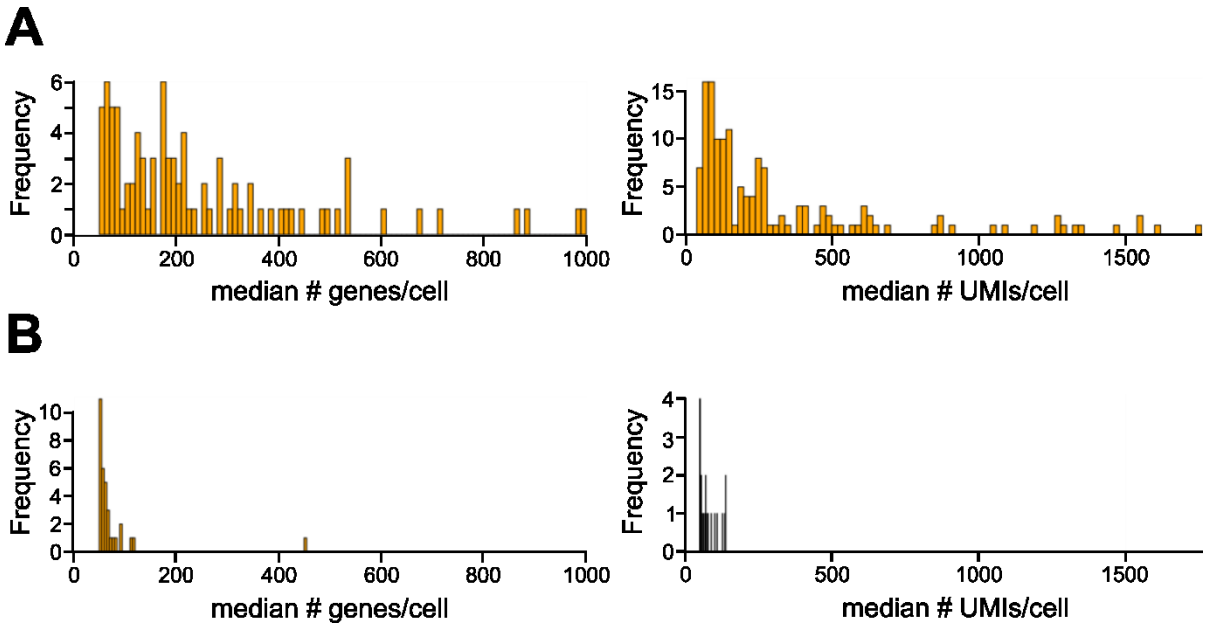


Figure 31: Methanol fixation allows more efficient recovery of genes and UMIs/cell than PBS/Triton X-100 permeabilization and para-formaldehyde fixation.
(A) – Median # of genes and UMIs/cell from methanol-fixed and -permeabilized HEK293 cells carried through the 3D-seq protocol. **(B)** - Median # of genes and UMIs/cell from PFA-fixed and PBS/Triton X-100-permeabilized HEK293 cells carried through the 3D-seq protocol.

Previously published combinatorial indexing protocols intent to increase the throughput of single cells per experiment (Cao et al., 2017; Rosenberg et al., 2018). 3D-seq however uses combinatorial indexing for its versatility and is aiming to record the spatial information of each cell. This entails, that the tissue is first fixed and permeabilized, then the RT reaction is run *in situ* and afterwards the tissue is digested to obtain a single cell solution. In all other protocols the tissue is first digested and once a single cell solution is obtained, the cells are fixed and permeabilized which results in the loss of all spatial information.

Hence, it is crucial to evaluate whether methanol-fixed tissue, here adult mouse brain slices on standard microscope slides, can still be digested into a single cell solution after methanol fixation. As starting point the suggestions from the Worthington Tissue Dissociation Guide on digesting adult mouse brain were consulted (Worthington Biochemicals). Then, it was qualitatively evaluated whether papain, collagenase, TrypLE or Proteinase K were able to disintegrate methanol-fixed tissue (Figure 32 - A). Two fixation times, 10 and 30 min, were

tested. Optical inspection of the glass slide after 30 min at 37°C with the different enzymes showed, that Papain was most efficient in digesting the tissue, while Collagenase and Proteinase K only partly detached the tissue. TrypLE left the tissue almost completely intact. Still, PBS was used to collect the detached cells and tissue pieces into a tube to be able to evaluate the digestion further in a microscope. Papain showed by far the most complete digestion of the tissue into a single cell solution. This was improved by shaking the tissue during the Papain digest (Figure 32 - B). The measured cell sizes ranged from 6.14 µm to 17.92 µm, which is within the expected range of neural brain cells (Hamberger et al., 1971). This and additional TrypanBlue staining indicate that the visible round structures are indeed single cells or single nuclei.

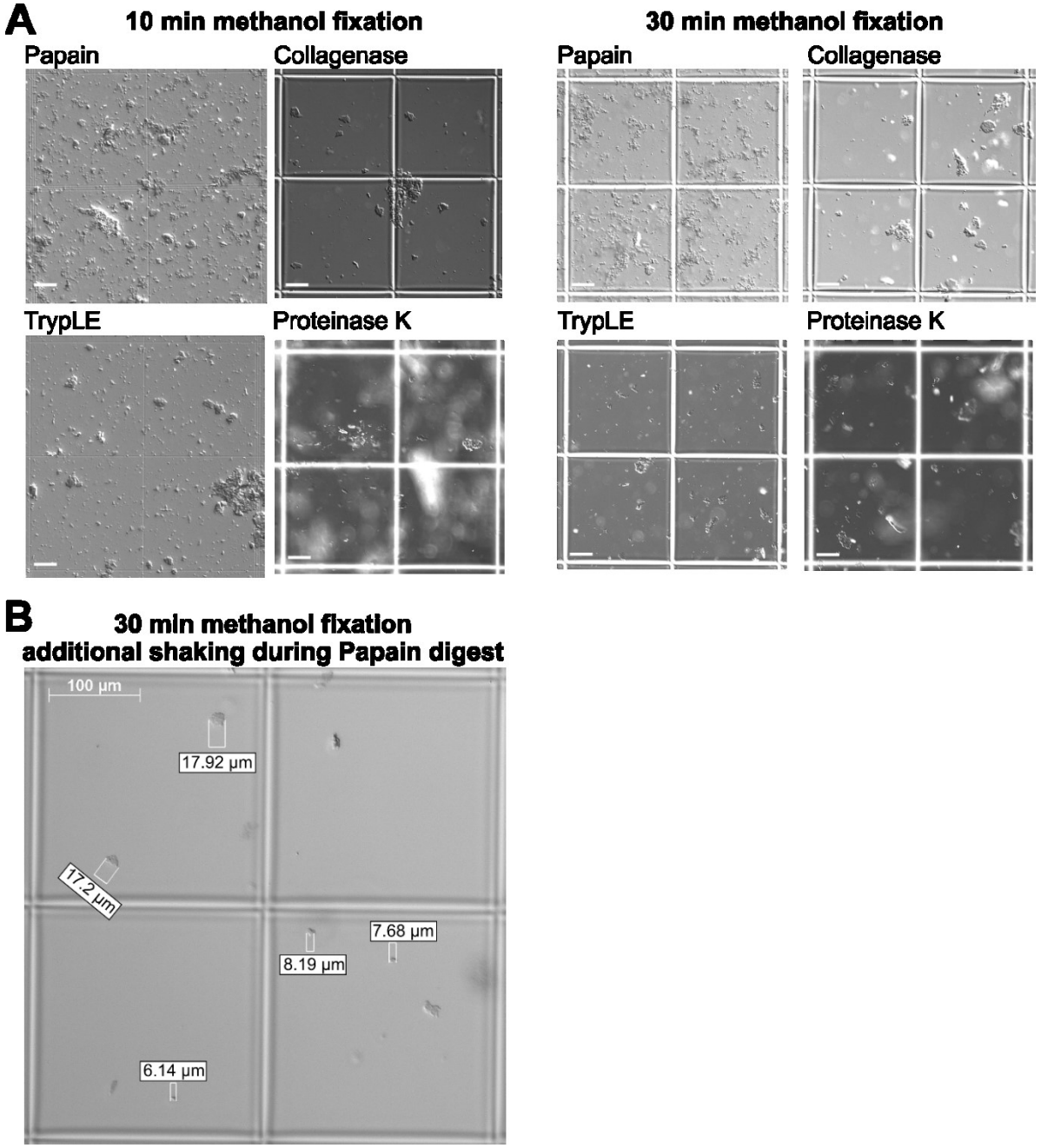


Figure 32: Methanol fixed mouse brain slices can be digested into single cells with Papain.
(A) – 10 min and 30 min methanol-fixed mouse brain slices were rehydrated and digested with papain, collagenase, TrypLE or Proteinase K (cf. 6.2.4 - Tissue digestion). Scale bar – 50 µm. **(B)** – Additional shaking during Papain digestion improves the retrieval of single cells further. Cell sizes were measured with AxioVision 4.8.

7.3.4 Optimization of material recovery throughout ligation, bead purification and cell pelleting

All single-cell sequencing protocols have to be designed to tackle very low input amounts. Hence, material loss throughout the individual steps of these procedures are a challenge that needs to be addressed. During 3D-seq, material loss can occur, next to the already mentioned reverse transcription step, also during ligation, during the cell pelleting steps and during bead purification (cf. Figure 29 – B).

During the ligation step a second barcode is added to the reverse transcribed molecules. At this step, the cells are still intact. First, it was analyzed whether changing certain parameters of the ligation step could increase its yield: adapter concentration (2 or 5 µM), reaction volume and thereby changing the enzyme concentration (10, 20 or 50 µl), different ligation enzymes (Figure 33 – A) by qPCR. The primer combination named “Ligation amount” (Figure 33) was used in these reactions. It was revealed that increasing the adapter concentration while keeping the reaction volume minimal increased the ligation yield 2-3-fold for T4 DNA Ligase. The Quick Ligation mix worked best for 2 µM adapter in a 10 µl reaction, the Blunt/TA ligation mix with 5 µM adapter in a 50 µl reaction. Considering the cost per reaction (per 96-well plate: 20,89 € (T4 DNA Ligase), 253,44 € (Quick Ligation mix), 152,06 € (Blunt/TA cloning) it was decided to further on use T4 DNA Ligase with 5 µM adapter in a 10 µl reaction for the 3D-seq protocol. Next, it was checked whether certain additives (PEG8000, 20 U/µl T4 DNA Ligase instead of 10 U/µl and 10 µM adapter instead of 5 µM) can increase the ligation efficiency further (Figure 33 – B). For this, qPCR was used with two sets of primers: one that measures all reverse transcribed cDNA (“RT amount”) and one that measures all reverse transcribed and ligated cDNA (“Ligation amount”) (Figure 33). Since the two products differ by up to 50 bp, their amplification efficiency was determined by qPCR on a dilution series (data not shown) and included in the calculation of ligated RT product. Under standard conditions (5 µM adapter, 10 U/µl T4 DNA Ligase, 10 µl reaction) approximately 50 % of the reverse transcribed cDNA were also ligated with a second barcode. Addition of 4.2 % PEG8000 doubled the ligation efficiency for certain genes, while even higher concentrations of T4 DNA Ligase or of the adapter did not significantly enhance the ligation efficiency. Next, the influence of PEG8000 addition was investigated on transcriptome-wide level by next-generation sequencing to rule out any undesired effects. Two runs of 3D-seq were conducted with and without PEG8000 addition, one with tissue and one with a cell suspension. Surprisingly, many incomplete barcode combinations were found after addition of PEG8000, meaning that the second barcode was ligated unspecifically to not correctly reverse transcribed cDNA (Figure 33 – C). Thus, PEG8000 was not added in subsequent experiments as it increased the number of unspecific ligations.

In order to trace back as many cells of the intact tissue as possible within the 3D-seq grid *in silico*, as little cells as possible should be lost throughout any steps of the 3D-seq protocol. Most prone for cell loss are the cell mixing, washing and pelleting steps after reverse transcription and ligation. Hence, a number of pelleting conditions were analyzed by counting the cells before and after centrifugation at 4°C: increasing speed, increasing centrifugation time, using special coated tubes (Figure 33 – D). MAXYMUM recovery tubes used at 2500xg

for 10 min turned out to be most efficient in pelleting fixed cells as this setup recovered 59-63 % of cells.

AMPure XP bead purification is used in this protocol after cell lysis (cf. Figure 29 – B). To efficiently lyse the cells, different lysis buffers were tested, each of them use high detergent concentrations (0.2 % Sarkosyl, 0.15 % Triton X-100 or 0.25 % IGEPAL). The presence of detergents can disrupt the binding of DNA to the carboxylated beads even in the presence of PEG and salts. qPCR showed that 80.1 % of cDNA without the addition of any detergents can be recovered with AMPure XP beads, while only 51 % can be recovered in the presence of detergents (Drop-seq buffer with 0.2 % Sarkosyl), 68.9 % (0.15 %Triton X-100) or 59.1 % (0.25 % IGEPAL). To account for this effect in subsequent 3D-seq runs, solutions containing high detergent concentrations were diluted 4x with water before proceeding with AMPure XP bead purification.

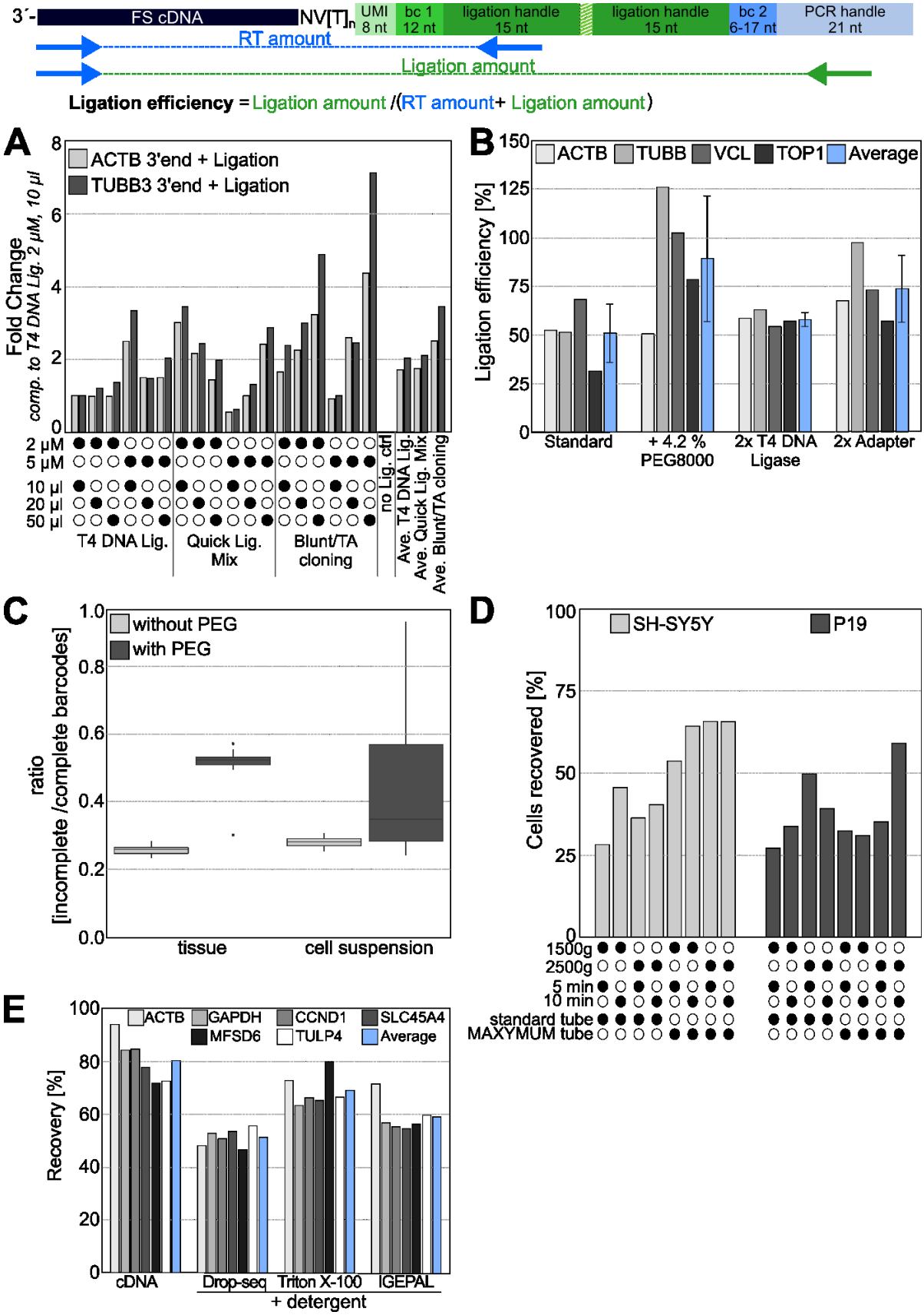


Figure 33: Minimizing material loss in ligation, cell pelleting and bead purification

(A) – Comparison of several ligation conditions, changed adapter concentration, changed reaction volume, several ligation enzyme mixes. **(B)** – qPCR quantification of influence from certain ligation reaction additives. **(C)** – Ratio of incomplete/complete barcodes quantified from sequencing for a 3D-seq run on tissue and on cell suspension with and without addition of PEG8000 to the ligation reaction. **(D)** – Optimization of cell recovery after pelleting. **(E)** – Percentage of recovered cDNA after AMPure XP bead purification of cDNA without detergents present, cDNA in Drop-seq buffer (0.2 % Sarkosyl), Triton X-100 lysis buffer (0.15 % Triton X-100) and IGEPAL lysis buffer (0.25 % IGEPAL).

7.3.5 Designing and producing a grid with CAD software and a 3D printer

A central part of the 3D-seq approach is to compartmentalize the tissue with a grid. Using a grid has several advantages, e.g. it is simple to use, several grids with different geometries can be applied consecutively to increase the final resolution and the compartmentalization with a grid will leave the tissue intact. A CAD software, Autodesk Fusion 360, was used to design a grid (Figure 34 – A) and a fitting mounting device (Figure 34 – B). The mounting device has recesses that hold the grid in place, when it is put onto a tissue on a microscope slide. Furthermore, the mounting device has drill holes for 3.5 screws that are then used to fasten the grid onto the tissue and minimize leakiness (Figure 34 – C). This grid design was then used to be printed on an ultimaker 2+ with polyamide (PLA) or polycarbonate (PC) as printing material. PLA is a more flexible and easier to print material that allows to build finer structures, while PC is a tougher material that bends less and has higher heat resistance. The PLA available for the ultimaker 2+ can only withstand temperatures up to 40°C (= glass transition temperature, engineering term referring to the temperature at which a hard material becomes brittle and then viscous/rubbery). Hence, it started bending and melting during standard reverse transcription conditions and was therefore incompatible with the 3D-seq protocol (data not shown). Despite PC is solid up to 147°C (= glass transition temperature) it was not suitable for the 3D-seq setup either, as it was impossible to print fine grid walls of the necessary thickness (100 – 300 µm) using the ultimaker 2+ (data not shown). Thus, a Projet HD 3000 was used with a proprietary PA-mixture (VisiJet SR200) that was able to print grids that both were sufficiently heat stable (glass transition temperature 46°C, (Vargas et al., 2014)) and contained finer walls (100 – 300 µm) (Figure 34 – D). An important parameter for the leakiness of the grid is the wall thickness, the thicker the wall the less likely are leaks caused by imperfect walls. However, increasing the thickness also increases the amount of tissue lost for analysis as it is crushed underneath the grid walls. Thus, grids of several wall thicknesses were printed and microscopically analyzed (Figure 34 – D). While the 100 µm walls exhibited a number of imperfections, the 300 µm walls were more even, just like the 150 and 200 µm grids (data not shown). Therefore, the 150 µm grid was the preferred choice to minimize material loss. When the grid is applied to an unstained mouse brain section (Figure 34 – F), anatomical structures of the brain like the cortex, the hippocampus, the corpus callosum, and so on, are still visible and help the orientation in later analysis. Next, the leakiness of the grid was investigated by plating cell lines from mouse (P19) and human (SH-SY5Y) in a predefined pattern onto a PLL-coated microscope slide with the help of an ibidi cell chamber (Figure 34 – E). Then, the cells were fixed in methanol, the 300 µm grid was applied, and the 3D-seq protocol was performed. The position of the plated cell lines under the grid was known and used to compare the obtained next-generation sequencing data to the expected cell identities (Figure 34 – E). Of the sequenced cells only 151 had more than 4 UMIs identified and were considered in this analysis. The identity of the cell was determined by the number of UMIs mapping unambiguously to a certain species – mouse or human. If a cell contained more than 20 % UMIs from both species, it was called a “mixed” cell. Such mixed cells can either occur due to leakiness of the grid, as the barcode combinations would not be sufficient anymore to identify single cells, or due to mixing of cell content later during the protocol. It was shown in previous combinatorial indexing experiments that the mixing during later protocol steps is minimal (Rosenberg et al., 2018), hence the mixing rate estimated here should be due to the leakiness

of the grid. 145 out of 151 considered cells have 100 % species purity (< 20 % UMIs from the other species), while 6 out of 151 cells were mixed (≥ 20 % UMIs from the other species). 6 mixed cells out of 151 total cells would correspond to 3.97 % mixed cells.

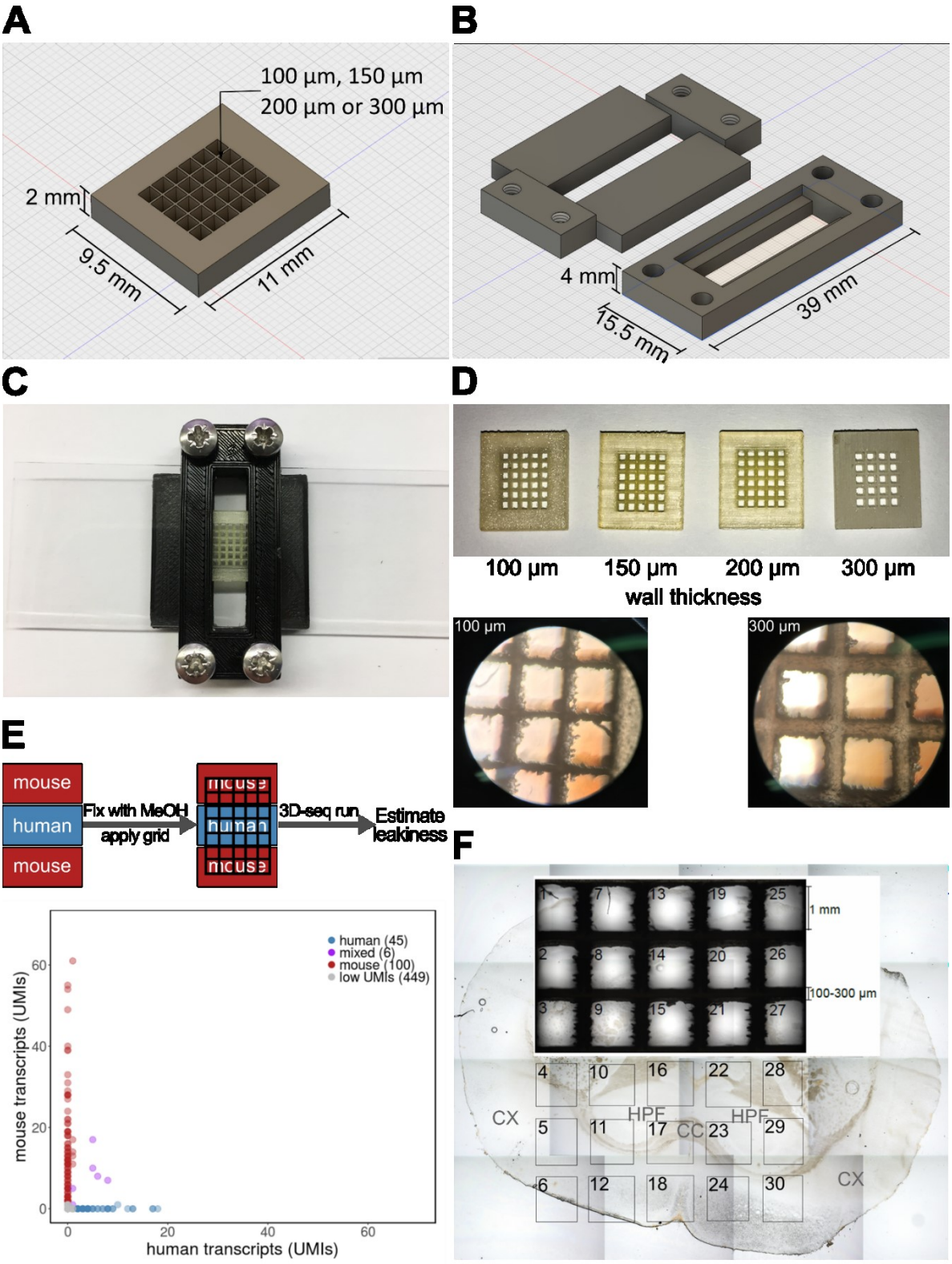


Figure 34: Designing, producing and evaluating the grid.
(A) – Image of grid from Autodesk Fusion 360. **(B)** – Image of mounting device, left: bottom part; right: top part. **(C)** – mounting device with grid inside. **(D)** – Grids produced with Projet HD 3000. **(E)** – Estimation of leakiness. Cells are seeded on a microscope slide in a known pattern, fixed with methanol and then the grid is applied. A 3D-seq run is conducted and the leakiness is estimated *in silico* by comparing the expected organism identity to the *in silico* identified one. **(F)** – Grid when applied to a coronal mouse brain section. CX – cortex, CC – Corpus callosum, HPF – Hippocampal formation.

7.3.6 3D-seq allows the identification of cell types from adult mouse brain slices

3D-seq is supposed to be applied to tissue sections that are placed on standard microscope glass slides. This has several advantages, first 3D-seq is compatible with already existing tissue cuts e.g. from precious medical samples, second standard sectioning protocols exist for almost all tissues, third no specialized equipment is needed for the initial preparation step of the tissue and fourth a microscopic inspection of the tissue with high resolution is possible and enables the collection of additional information about the tissue. To develop the method, we are applying it to coronal sections from adult mouse brain (Figure 35 - A). Since anatomical structures are visible within the tissue even without additional staining, they can manually be annotated with the help of the Allen Brain Atlas (Allen Institute for Brain Science, 2011) and later on be used as additional information.

The shown mouse brain section was then carried through the 3D-seq protocol, sequenced and yielded 80 UMIs/cell (median) or 158.8 UMIs/cell (mean) and 16 genes/cell (median) and 27.2 genes/cell (mean) when considering all cells with more than 50 UMIs expressed and counting only genes with at least 2 UMI counts. The data obtained here, as for any single-cell RNA-sequencing experiment, is of high dimensionality. Hence, a nonlinear dimensionality reduction method, the t-distributed stochastic neighbor embedding (tSNE), that was previously suggested for dimensionality reduction (Amir et al., 2013; van der Maaten and Hinton, 2008), was used. First, a principal component analysis was conducted to determine the statistically significant principal components (PCs), 100 PCs were then reduced to two dimensions using tSNE (Figure 35 - C). 71 clusters were identified with 50 (cluster 70) to 584 cells (cluster 0) per cluster (Figure 35 – D). The number of UMIs/cell is equally distributed across the clusters, indicating that the cluster identity is not build on noise in the data but on actual biological meaning. However, it should be noted that with deeper data, e.g. recovering more UMIs or genes/cell some clusters will collapse and be united with others as the sparse data can let cells seem different on the transcriptome level. Hence, marker genes for each cluster were determined and biological identities assigned with the help of mousebrain.org (Linnarson lab) (Table 7).

Table 7: Marker genes identified for clusters from coronal mouse brain section

cluster	# of cells	gene	p-value[1]	Ave. logFC[2]	Positive in cluster[3]	Positive in all[4]	Adj. p-value[5]	cell type mousebrain.org	brain region
4	146	Zdhhc14	0	4.65	0.44	0.00	0	Excitatory neurons, mostly cerebral cortex & HC	Cortex, HC & minor in other regions
4	146	Myo16	0	4.40	0.31	0.00	0	Excitatory neurons, mostly cerebral cortex & HC	Cortex, HC & minor in other regions
4	146	Osbpl8	8.65E-247	4.26	0.31	0.01	1.29E-242	Excitatory neurons, mostly cerebral cortex & HC	Cortex, HC & minor in other regions
5	133	Gad2	0	4.18	0.29	0.00	0	Trilaminar cells, hippocampus	Cortex/HC thalamus
6	124	Snap25	3.62E-287	3.51	0.59	0.02	5.38E-283	Excitatory neurons, cerebral cortex	Cortex/HC
7	122	Ubr5	0	4.23	0.40	0.01	0	R-LM border Cck interneurons, cortex/hippocampus	Cortex/HC

7	122	Ahi1	0	3.94	0.40	0.00	0	R-LM border Cck interneurons, cortex/hippocampus	Cortex/HC
7	122	Adarb2	4.21E-119	3.59	0.28	0.01	6.25E-115	R-LM border Cck interneurons, cortex/hippocampus	Cortex/HC
8	114	Cadps2	0	4.37	0.40	0.00	0	all kinds of neurons	everywhere
8	114	Ankrd17	0	4.01	0.44	0.01	0	all kinds of neurons	everywhere
9	107	Sparcl1	0	4.04	0.73	0.02	0	inhibitory and excitatory neurons	
9	107	Abca3	3.27E-269	3.49	0.30	0.00	4.87E-265	inhibitory and excitatory neurons	
10	104	Dst	0	4.36	0.81	0.02	0	Excitatory neurons, cerebral cortex	Cortex/HC
11	102	Herc1	0	4.33	0.53	0.01	0	all kinds of neurons	everywhere
12	101	Kif5c	0	3.98	0.90	0.02	0	all kinds of neurons	everywhere
13	101	Gls	0	4.98	0.54	0.00	0	Cholinergic neurons, septal nucleus, Meissnert and diagonal band	Hypothalamus
13	101	Rimbp2	0	4.95	0.50	0.00	0	Cholinergic neurons, septal nucleus, Meissnert and diagonal band	Hypothalamus
14	100	Cdh13	0	4.76	0.69	0.01	0	Excitatory neurons, cerebral cortex or midbrain	cortex, midbrain
14	100	Pex5l	1.25E-251	4.49	0.34	0.01	1.86E-247	Excitatory neurons, cerebral cortex or midbrain	cortex, midbrain
15	100	Tulp4	0	4.76	0.59	0.00	0	Non-border Cck interneurons, cortex/hippocampus	Cortex/HC
15	100	Ralgapa2	0	4.01	0.36	0.00	0	Non-border Cck interneurons, cortex/hippocampus	Cortex/HC
16	99	Cdk17	0	4.85	0.52	0.00	0	Excitatory neurons, cerebral cortex	Cortex/HC
16	99	Gm15800	0	4.68	0.51	0.00	0	Excitatory neurons, cerebral cortex	Cortex/HC
17	95	Cplx2	0	4.37	0.84	0.01	0	all kinds of neurons	everywhere
18	94	Susd4	0	4.79	0.44	0.00	0	Excitatory neurons	Cortex, HC, thalamus, hypothalamus
18	94	Nin	0	4.63	0.59	0.00	0	Excitatory neurons	Cortex, HC, thalamus, hypothalamus
19	93	Map1b	0	3.56	1.00	0.04	0	Afferent nuclei of cranial nerves VI-XII	Medulla
20	93	Grin2b	0	3.39	1.00	0.05	0	Excitatory neurons, cerebral cortex	Cortex/HC
21	92	Rgs7	0	4.39	0.77	0.01	0	Excitatory neurons, cerebral cortex	Cortex/HC

21	92	Cltc	7.91E-159	3.26	0.26	0.01	1.18E-154	Excitatory neurons, cerebral cortex	Cortex/HC
22	89	Ptprg	0	5.15	0.87	0.01	0	Excitatory neurons, cerebral cortex	
23	89	Kif5a	0	3.34	1.00	0.05	0	Excitatory neurons, cerebral cortex	
24	89	Fam19a2	0	5.27	1.00	0.01	0	Excitatory neurons, cerebral cortex	
25	88	Cacna2d3	1.09E-261	3.01	0.94	0.05	1.62E-257	Non-border Cck interneurons, cortex/hippocampus	Cortex/HC
26	87	Cadps	0	4.38	1.00	0.02	0	Excitatory neurons, cerebral cortex	Cortex/HC
27	86	Sntg1	0	4.38	1.00	0.02	0	Excitatory neurons, cerebral cortex	Cortex/HC
28	85	Kif5b	0	5.56	1.00	0.00	0	Excitatory neurons, cerebral cortex	Cortex/HC
29	85	Tcf4	0	4.18	1.00	0.02	0	Excitatory neurons, cerebral cortex and oligodendrocytes	Cortex/HC, thalamus
30	85	Gpr158	0	4.45	1.00	0.02	0	Excitatory neurons, cerebral cortex	Cortex/HC, thalamus
31	83	Mbp	0	3.59	1.00	0.04	0	oligodendrocytes	everywhere
32	81	Calm1	0	3.66	1.00	0.03	0	Excitatory neurons	Cortex/HC, thalamus
33	79	Syne1	0	3.91	1.00	0.03	0	Excitatory neurons, cerebral cortex	Cortex/HC
34	78	Cacna1c	0	4.31	1.00	0.02	0	Excitatory neurons	Cortex, HC, thalamus, hypothalamus
35	78	Wwox	0	5.22	1.00	0.00	0	Excitatory neurons, cerebral cortex	Cortex/HC
36	78	Limch1	0	5.46	1.00	0.00	0	Excitatory neurons	Cortex, HC, thalamus, hypothalamus
37	77	Dclk1	0	4.79	1.00	0.01	0	Excitatory neurons, cerebral cortex	Cortex/HC
38	77	Mbnl2	4.05E-263	4.47	0.47	0.01	6.03E-259	Afferent nuclei of cranial nerves III-V	hypothalamus
38	77	Frmd4b	2.10E-175	3.97	0.33	0.01	3.13E-171	Afferent nuclei of cranial nerves III-V	hypothalamus
38	77	Lmtk2	1.48E-139	3.11	0.26	0.01	2.20E-135	Afferent nuclei of cranial nerves III-V	hypothalamus
39	76	Ppp2r2c	0	5.17	1.00	0.01	0	Excitatory neurons	everywhere
40	76	Cacna1d	0	5.33	1.00	0.01	0	Excitatory neurons, midbrain	Cortex, Thalamus
41	76	Alcam	0	5.09	1.00	0.01	0	broadly expressed	everywhere
42	76	Cmip	0	5.70	1.00	0.00	0	Excitatory neurons	everywhere
43	75	Kcnq3	0	4.36	1.00	0.02	0	Excitatory neurons	everywhere
44	75	Cdc42bpa	0	4.81	1.00	0.01	0	Excitatory neurons	everywhere

45	73	Ptms	0	5.45	1.00	0.00	0	broadly expressed, highest in enteric glia	everywhere
46	71	Camk2n1	0	4.19	1.00	0.03	0	Excitatory neurons, cerebral cortex	Cortex
47	70	Elmo1	0	5.34	1.00	0.01	0	Non-border Cck interneurons	Cortex/HC
48	69	Unc79	0	5.41	1.00	0.00	0	Excitatory neurons, cerebral cortex	Cortex/HC
49	69	Ttc3	0	3.86	1.00	0.03	0	Excitatory neurons	everywhere
50	68	Dnm1	0	5.32	1.00	0.01	0	Excitatory neurons, cerebral cortex	Cortex/HC
51	68	Kif1b	0	3.50	1.00	0.03	0	Excitatory neurons, cerebral cortex	Cortex/HC
52	68	Slc8a1	0	5.03	1.00	0.01	0	broadly expressed	everywhere
53	67	Frrs1l	0	5.18	1.00	0.01	0	Excitatory neurons	everywhere
54	67	AI314180	0	5.65	1.00	0.00	0		
55	67	Nell1	0	5.44	1.00	0.00	0	broadly expressed, highest in inhibitory interneurons of HC	everywhere
56	63	Atp2a2	0	5.45	1.00	0.00	0	Afferent nuclei of cranial nerves VI-XII	medulla, hypothalamus
57	63	Nfix	0	5.56	1.00	0.00	0	broadly expressed, highest in granule neuroblasts	dentate gyrus, cortex
58	63	Rbfox2	0	5.13	1.00	0.01	0	Excitatory neurons	everywhere
59	63	Rnf150	0	4.73	1.00	0.01	0	Excitatory neurons	everywhere
60	62	Npas3	0	5.29	1.00	0.01	0	(Non-border Cck) interneurons	Cortex/HC
61	62	Ptprj	0	5.62	1.00	0.00	0	Excitatory neurons	Cortex/HC
62	62	Peg3	0	5.34	1.00	0.00	0	serotonergic neurons	hypothalamus
63	62	Sh3rf3	0	5.63	1.00	0.00	0	Non-border Cck interneurons	Cortex/HC
64	61	Calm3	0	5.57	1.00	0.00	0	broadly expressed, neurons	everywhere
65	60	Kcnh1	0	5.35	1.00	0.00	0	R-LM border Cck interneurons	Cortex/HC
66	60	Snrnp70	0	5.09	1.00	0.00	0	broadly expressed, all cell types	everywhere
67	58	Ptprn2	0	4.00	1.00	0.02	0	broadly expressed, neurons	everywhere
68	55	Diaph2	0	5.36	1.00	0.01	0	broadly expressed, neurons	everywhere
69	55	Plxna4	0	5.16	1.00	0.01	0	excitatory neurons	everywhere
70	50	Kcnip1	0	5.39	1.00	0.00	0	broadly expressed, neurons	Cortex/HC, thalamus

[1]calculated with Wilcoxon rank sum test; [2]average logFC in cluster vs. all cells; [3] positive cells in cluster [%/100]; [4]positive cells of all cells [%/100]; [5]p-value adjusted with Bonferroni method

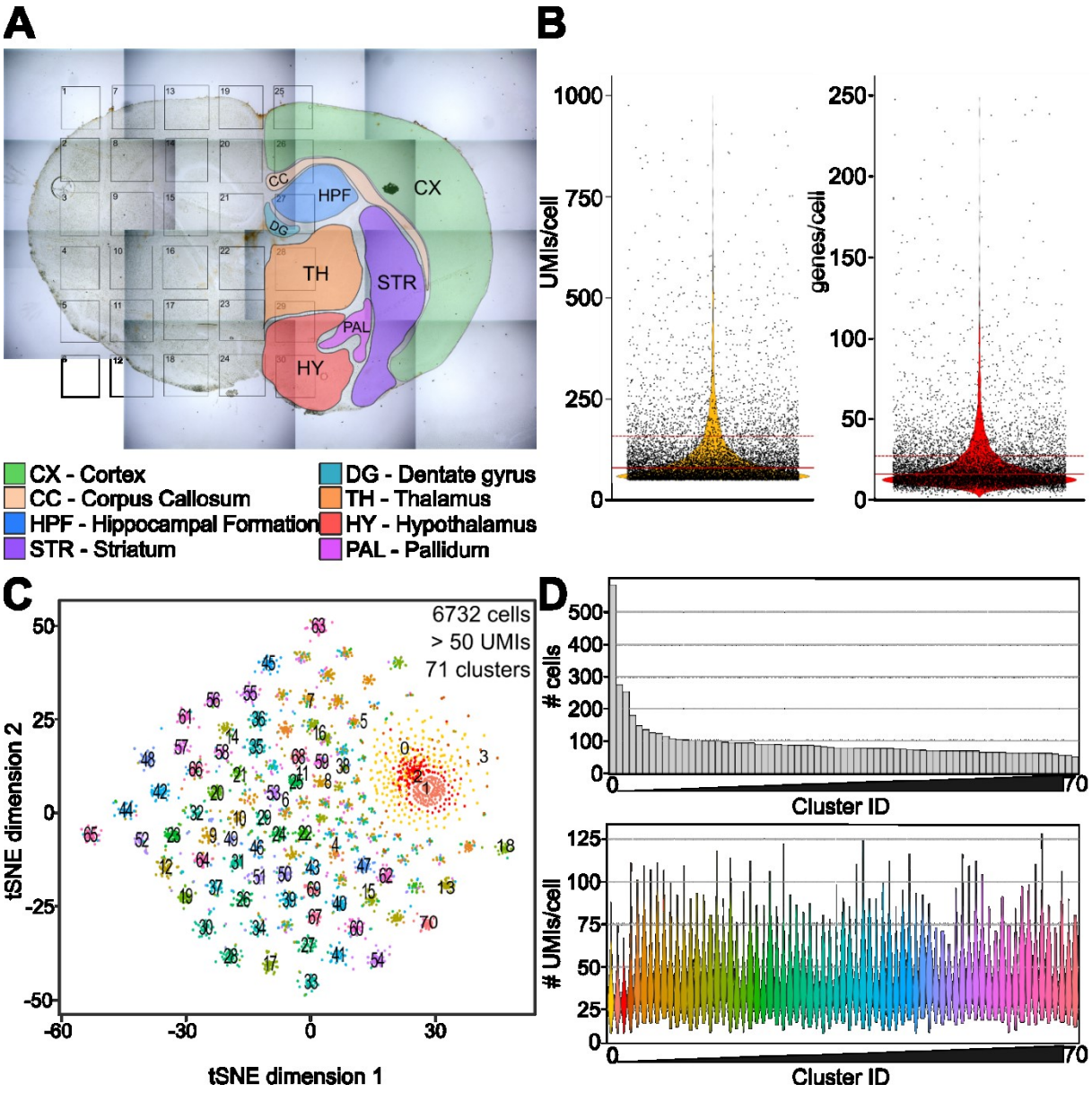


Figure 35: 3D-seq run on coronal mouse brain section allows identification of cell types.
(A) – Stitched microscopic images of a coronal mouse brain section. Positions of grid compartments (1x1 mm) are annotated, as well as major mouse brain regions according to the Allen Mouse Brain Reference Atlas. CX – Cortex, CC – Corpus Callosum, HPF – Hippocampal Formation, STR – Striatum, DG – Dentate Gyrus, TH – Thalamus, HY – Hypothalamus, PAL – Pallidum. **(B)** – Violin plots for UMIs/cell (only cells with > 50 UMIs were kept for this analysis) and genes/cell. Each dot is a cell. Horizontal line = median, dashed horizontal line = mean. **(C)** – tSNE plot for 6732 cells from a coronal mouse brain section carried through the 3D-seq protocol with more than 50 UMIs yield 71 clusters, based on 100 PCs, resolution = 3, perplexity = 30. **(D)** – Number of cells per cluster and number of UMIs/cell in each cluster.

7.3.7 Spatial reconstruction of mouse brain *in silico*

The main aim of 3D-seq is the unbiased spatial reconstruction of tissue *in silico*. During the RT barcoding step all cells are labeled with a specific RT barcode depending on their localization within the grid. This barcode can then be used for tracing the cells back to their approximate original position within the tissue. The resolution is limited to the size of each grid compartment. Here, this would mean that each cell is placed back within 1 mm^2 of its original position.

First, the distribution of UMIs/cell and genes/cell across the tissue was analyzed to rule out grid induced artifacts, e.g. that certain grid compartments contribute cells with high UMIs while others contribute only cells with low UMIs (Figure 36). Notably, two compartments are almost devoid of cells. The upper left corner compartment did contain only a small tissue part (cf. Figure 35 – A) which could explain that there was only one cell detected. However, the 4^{th} compartment in the 4^{th} column is also empty although it contained large parts of the thalamus (cf. Figure 35 – A). This is probably due to a hydrophobic effect: the grid is made from hydrophobic material while the reverse transcription mixture is water-based. Hence, if the drop of RT mixture is not placed right on the tissue the mixture will be held above the tissue. Nevertheless, all other compartments hold 1 - 787 cells with varying numbers of UMIs and genes. Hence, the number of UMIs/cell or genes/cell does not dependent on the position of the cells in the grid.

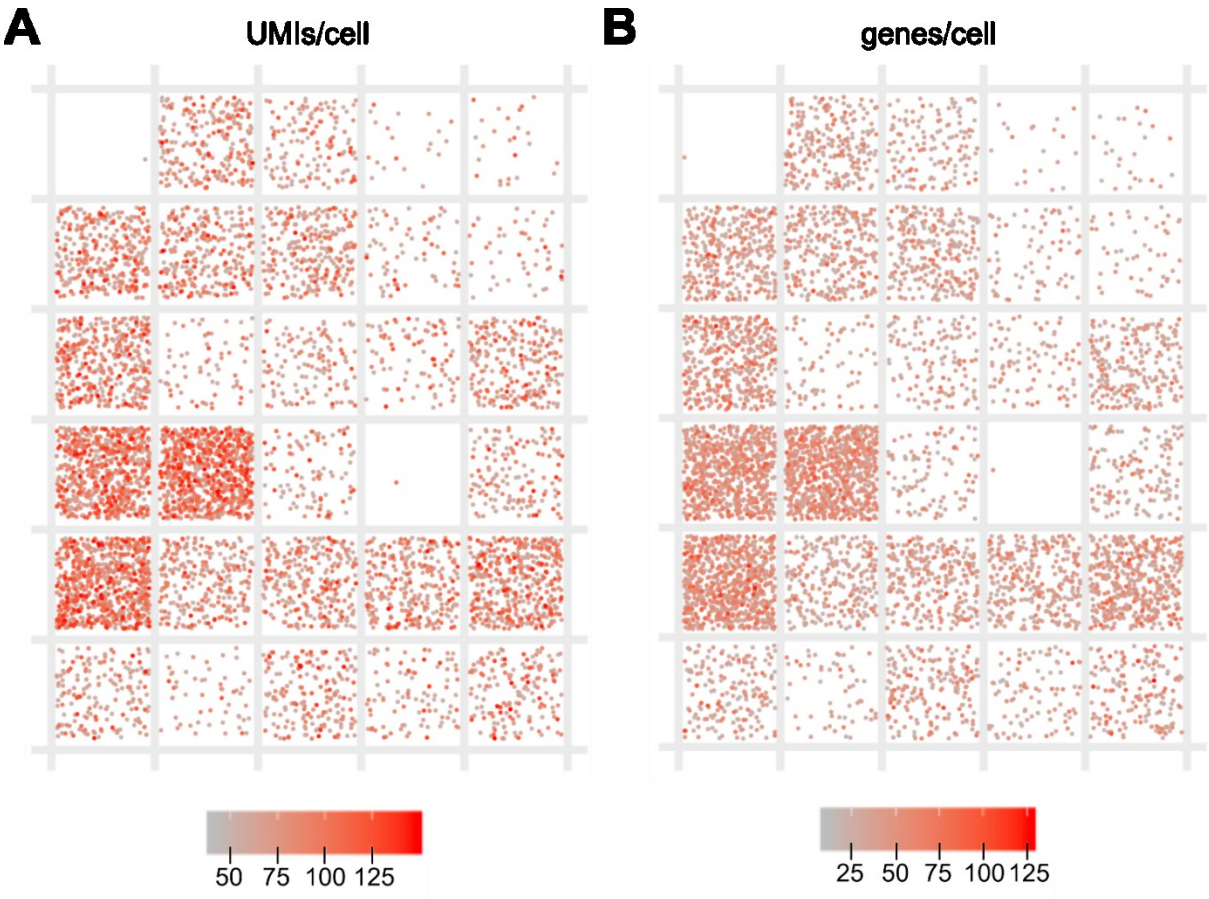


Figure 36: Distribution of UMIs/cell and genes/cell in each compartment of the grid.
(A) – UMIs/cell in each compartment. Cells (dots) are colored by number of UMIs detected. **(B)** – Genes/cell in each compartment. Cells (dots) are colored by number of genes detected.

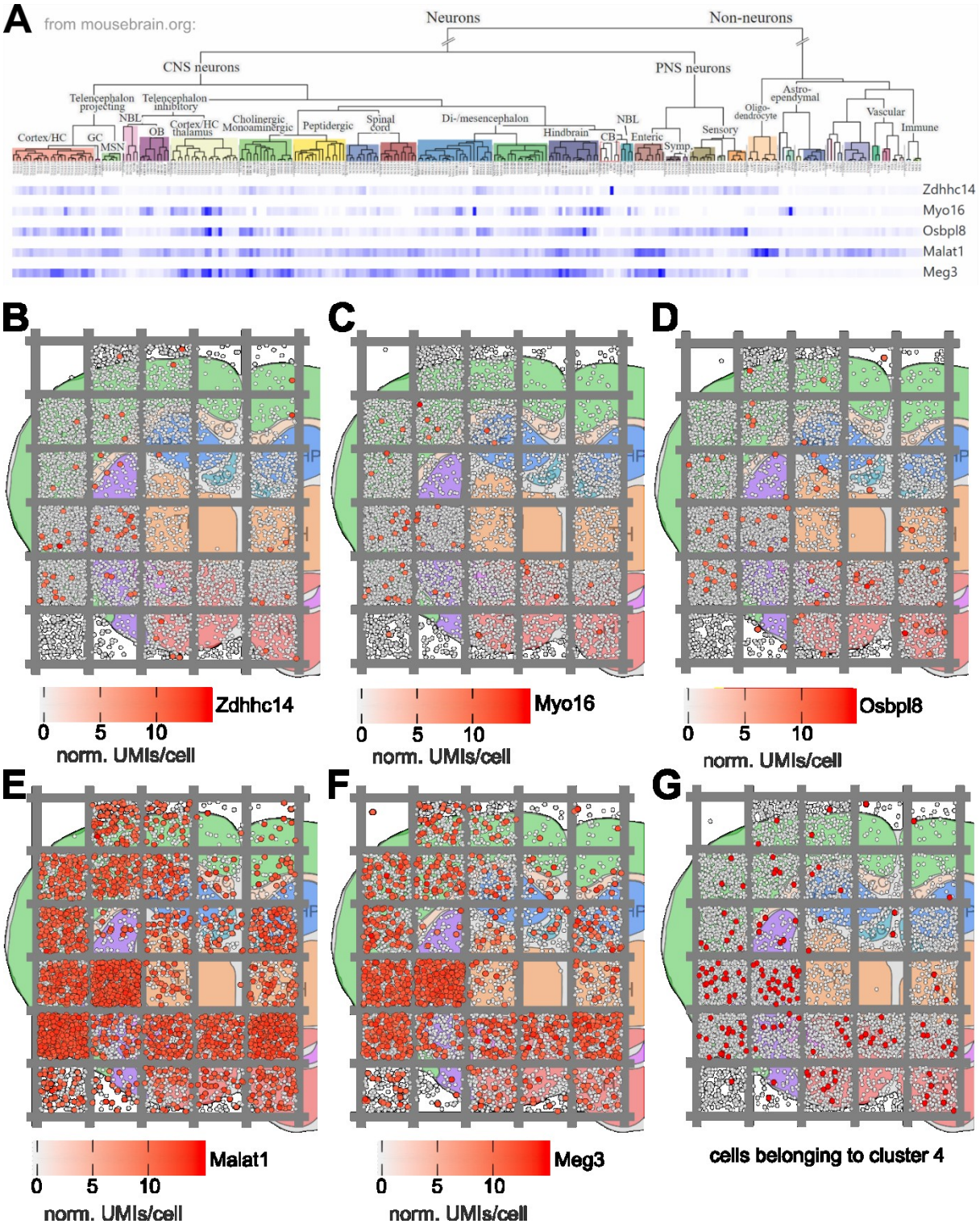


Figure 37: 3D-seq allows *in silico* reproduction of a spatial gene expression map.
(A) – Image from mousebrain.org showing the expression of Zdhhc14, Myo16 and Osbpl8 (marker genes for cluster 4) as well as Malat1 and Meg3 (for comparison). **(B, C, D, E, and F)** – Spatial gene expression maps based on 3D-seq data. Each dot is a single cell, grey – no UMIs for gene detected, red – cells with UMIs for respective gene detected and colored according to expression level. **(B)** – Zdhhc14. **(C)** – Myo16. **(D)** – Osbpl8. **(E)** – Malat1. **(F)** – Meg3. **(G)** – Spatial mapping of all cells belonging to cluster 4 based on 3D-seq data. Grey – cells of other cell clusters, red – cells belonging to clusters.

The mouse brain is a very well analyzed tissue in terms of spatial gene localization, resources like mousebrain.org (Linnarson lab) and the Allen Brain Atlas (Allen Institute for Brain Science, 2011) offer large single-cell RNA-seq datasets that are prepared from specific brain regions or *in situ* hybridizations of a wide variety of genes, respectively. Mousebrain.org additionally offers a quantitative measure for gene expression per cell type in certain brain regions (Figure 37 - A). In our data set Zdhhc14, Myo16 and Osbpl8 were identified as marker genes for cluster 4. Analyzing the individual gene expression patterns for these genes reveals that they are usually expressed in more than one cell type. Zdhhc14 is expressed in a number of neurons of the central nervous system (CNS), mostly in cortical and hippocampal neurons, and can also be found in non-neuronal cells, e.g. in oligodendrocytes. Evaluating the spatial gene expression of Zdhhc14, obtained by 3D-seq, confirms this. Most Zdhhc14 positive cells are found in the cortex, but also other regions such as the hypothalamus and striatum hold Zdhhc14-positive cells. A similar pattern is expected and found for Myo16. Osbpl8 on the other hand is higher and more broadly expressed than Zdhhc14 and Myo16, according to mousebrain.org. Hence, for Osbpl8 there are also cells found in the hippocampal area and in the thalamus (Figure 37 – B, C, and D). However, the data analyzed is sparse (cf. Figure 35 – B and D). Therefore, it will contain false negatives. According to the 3D-seq dataset 70 cells express Zdhhc14 (1.04 % of all cells), 53 cells express Myo16 (0.78 %), and 80 cells express Osbpl8 (1.19 %). For comparison, the spatial patterns of two highly expressed genes, Malat1 which is expressed in almost all cells, and, Meg3 which is expressed in all neuronal cells (Linnarson lab). In the 3D-seq dataset Malat1 is expressed in 2411 (35.81 %) cells and Meg3 in 1168 cells (17.35 %). So, even generally highly and widely expressed genes are only detected in one-fifth to one-third of the cells (Figure 37 – E and F). Therefore, the spatial expression trends recovered here are limited and some trends might not be detectable at this point. An alternative to displaying spatial expression on the gene level is the *in silico* assignment of cells belonging to a certain cluster to their original place within the tissue. The three marker genes for cell cluster 4, Zdhhc14, Myo16 and Osbpl8 are widely expressed, all three of them are moderately expressed in excitatory neurons of the cerebral cortex and hippocampus according to mousebrain.org. Hence cluster 4 was assigned the identity ‘excitatory neurons, mostly cerebral cortex and hippocampus’. Indeed, roughly half of the 146 cells belonging to cluster 4 are localized in grid compartments holding parts of the cortex (75 cells, 51.37 %). Further cells of cluster 4 are found in the hypothalamus, striatum and thalamus according to 3D-seq data. Mousebrain.org confirms this: at least 2 of the 3 markers are also expressed in cell types of the striatum (MSN in Figure 37 - A) or the di- and mesencephalon, which comprises the thalamus and hypothalamus.

8 Discussion

8.1 Circular RNAs in the Mammalian brain are highly abundant, conserved and dynamically expressed

In order to obtain a broad overview of circRNA expression in neural systems, a circRNA catalogue was created. To do so, 29 different types or stages of neural cell culture models and tissues were investigated by total RNA-seq. circRNA expression was found to be stage-specific and in many cases tissue specific. circRims2 for example is most highly expressed in the cerebellum, while the expression of the corresponding mRNA is highest in the striatum. This suggests a regulated, region-specific expression of circRNAs. Furthermore, circRNAs are strongly enriched in synaptoneurosomes compared to their linear counterparts. One explanation could be that circular RNAs diffuse to these distal cell parts and persist longer because of their exceptional stability. Another explanation for the enrichment of circRNAs at the synapse could be their active transport to this cell part. How mRNAs, RBPs or miRNAs are transported to the synapses it not fully understood (Dahm and Kiebler, 2005; Doyle and Kiebler, 2011; Holt and Bullock, 2009). However, all of these molecules are exerting specialized functions at the synapse that require their specific transport to this localization. Although there are no insights into the function of circRNAs at the synapse yet, this finding is quite intriguing and worth following up onto. Since circRNAs can be sliced open by microRNAs (Hansen et al., 2013), they could function as cargo-carrying adaptor molecules. Furthermore, when considering the long half-life of circRNAs compared to their linear counterparts (Zhang et al., 2016) their predominant localization at the synapse seems reasonable from an energetic point of view. Transporting an mRNA or a circRNA to the tip of an axon will cost a similar amount of ATP. Considering the longer half-lives of most circRNAs compared to the corresponding mRNA the transport needs to happen less frequently to maintain the same molecule concentration at the synapse. Synaptic density in human is approximately 4 times higher than in mouse brain (Herculano-Houzel, 2009), which could explain the generally higher number of circRNAs in human brain compared to mouse brain. Another difference between mouse and human is the expanded repertoire of inverted *Alu* repeats in primate introns (Daniel et al., 2014), that were already shown to be important in circRNA biogenesis (Ivanov et al., 2015).

circRNA expression was overall upregulated during neuronal differentiation. In both, cell lines (SH-SY5Y and P19) and primary neuron culture, in many cases but not always, the upregulation of the circRNA was coupled to the upregulation of the mRNA. This would argue that it is not the overall transcriptional output from a genetic locus that drives circRNA expression, but that production of mRNA or circRNA from a genetic locus can be specifically regulated. But how could the expression of the circRNA be specifically regulated? So far, the insights into the biogenesis of circRNAs are limited. However, splice regulators such muscleblind (MBL/MBNL) have been implicated in circRNA biogenesis (Ashwal-Fluss et al., 2014), as well as RBPs like QKI (Conn et al., 2015), ILF3 (Li et al., 2017), and the RNA-editing enzyme ADAR1 which antagonizes circRNA production (Ivanov et al., 2015). In human cell culture systems ADAR1 was also found to negatively regulate the expression of circRNAs, as its knockdown caused the upregulation of many circRNAs but not their linear counterparts.

Furthermore, an editing site in an intron of circRTN4 was identified that correlated well with the observed expression changes of circRTN4 during neuronal differentiation. The overall regulation of circRNA expression by ADAR activity was also uncovered in *Drosophila melanogaster*. Interestingly, ADAR1 is known to be highly active during development and in neurons (Wahlstedt et al., 2009) which could indicate a highly dynamic regulation of circRNA expression by ADAR1 in these situations.

The high conservation of circRNA expression was especially intriguing. Approximately 80 % of circRNAs expressed with at least 5 reads in mouse P19 cells have a homologous circRNA in human or shares at least the splice sites with a circRNA expressed in human. In addition, a higher conservation of 7-mers in circRNA encoding regions was found. This could indicate additional functions of the nucleotide sequence besides coding for the open-reading frame of the corresponding mRNA, like binding of miRNAs, RBPs or pairing with other long RNAs. It could also be a feature of exons that are enclosed by long introns, which is often the case for circRNA producing exons. Conserved circular RNAs were flanked more often by introns with RCMs than non-conserved ones, which would suggest a conserved mechanism for circRNA production. However, *Alu* elements, which are primate-specific and highly abundant repeat elements, have been implied in gene regulation in general and are often found in genomic regions with a high gene density (Deininger, 2011). In human there are approximately 306 *Alu* elements per megabase and at most a subset of them will be influencing circRNA expression.

In summary, a comprehensive circRNA catalogue for many neural tissues is provided which offers insights in the spatial localization of circRNAs as well as in the regulation of their expression during neuronal differentiation. The spatial localization however is coarse and limited to averages over entire brain regions or all synaptoneurosomes. Yet, starting from this catalogue one can select from a large set of interesting circular RNA molecules to analyze their potential functional roles in neuronal differentiation.

8.2 circSLC45A4 is required to keep neuronal cells in a progenitor state in cell culture systems and in the mammalian brain

In the present work circSLC45A4 was identified as the highest expressed isoform from its genetic locus in developing human cortex, at gestational week 22. At gestational week 22 neurogenesis in the cortex is almost completed, while neuronal migration and maturation are only beginning (Stagni et al., 2015). Additionally, we measured circSlc45a4 levels in the whole mouse brain and found its expression to be peaking at E16.5. A timepoint at which neurogenesis in the mouse brain is peaking in many mouse brain regions (Finlay and Darlington, 1995) and at which upper-layer neurons are formed in the cortex (Mihalas et al., 2016). This indicated that circSLC45A4 might play a role in neurogenesis. Hence, its expression levels were perturbed by siRNA-mediated RNAi (Elbashir et al., 2001a) in a widely-used human model system for neuronal differentiation (SH-SY5Y) and in the developing mouse cortex. In SH-SY5Y cells spontaneous neuronal differentiation was observed following circSLC45A4 knockdown. Induction of neuronal marker genes like TUBB3, ASCL1 and GABBR1 was detected by qPCR and in part by Western Blot. At the same time, knockdown of circSLC45A4 in SH-SY5Y cells caused a significant increase in neurite length (19.9 µm after circSLC45A4 KD, 12.0 µm after scramble KD).

For both mRNA Slc45a4 and circSlc45a4 knockdown a significant depletion of cells from the cortical plate of mouse embryonic cortices was observed. However, only after knockdown of circSlc45a4 cells were also depleted from the subventricular zone, the layer that contains basal progenitors. Indeed, basal progenitors were found to be significantly depleted after circSlc45a4 knockdown. Basal progenitors can be generated from apical progenitors or basal progenitors themselves. Basal progenitors also undergo symmetric, neurogenic divisions and thereby generate neurons in the developing cortex (Taverna et al., 2014). In addition to the depletion of the basal progenitor pool, we observed a slight but significant increase in apical progenitors, the number of neurons, however, was not affected.

circSLC45A4 and mRNA SLC45A4 originate from the same genetic locus. Hence, they are identical in their sequence except for the head-to-tail splice junction that is specific to the circular RNA. Therefore, it is important to ensure that the effects described for perturbation of the circRNA are indeed caused by the circRNA and not through secondary effects on the mRNA. siRNAs and shRNAs are prone to have off-target effects, either by downregulation of unintended targets, e.g. the corresponding mRNA, functioning as miRNAs, depleting the effective concentration of AGO, or simply by overloading Dicer (shRNAs only) and thereby impairing its function. To minimize this risk several precautions were taken: a) several independent siRNAs against the same target were used and only effects observed for each of these siRNAs were considered, b) siSPOTR (Boudreau et al., 2013) was used to minimize the chance of hitting unintended targets with potential miRNA seed sequences included in the siRNA, c) BLAST (Altschul et al., 1990) analysis of the siRNA sequences was used to ensure that no other RNAs will be targeted by the siRNA. Furthermore, for each experiment conducted the levels of the corresponding mRNA were monitored and found stable or minimally affected. Additionally, in mouse cortex depletion of the mRNA caused a different phenotype than depletion of the circRNA. Although we cannot rule out unspecific effects, the observed

phenotypes are attributed to the depletion of the circRNA and likely not caused by off-target effects or deregulation of the mRNA.

So how can the phenotypes observed in SH-SY5Y cells, spontaneous induction of neurogenesis, and mouse cortex, depletion of the basal progenitor pool, be reconciled (Figure 38 - A)? In both cases a depletion of a neuronal progenitor pool was observed: SH-SY5Y cells are already mainly neuroblasts that are spontaneously differentiating towards a neuronal phenotype after the depletion of circSLC45A4. In mouse cortex several lines of evidence showed the depletion of the basal progenitor pool already in the ventricular zone. In addition to the depletion of the basal progenitor pool, we observed a slight but significant increase in apical progenitors, the number of neurons, however, was not affected.

What causes this distortion in the apical to basal progenitor pool ratio? Here the increased number of Cajal-Retzius cells that was determined from the combined analysis of bulk RNA-seq and single-cell RNA-seq data seems intriguing. Cajal-Retzius cells are the first type of neuron appearing in the developing neocortex. Their main function is to secrete reelin, an extracellular matrix glycoprotein, that helps to mediate neuronal migration and controls cell-cell interactions in the developing brain (Gupta et al., 2002). Furthermore, it was shown that reelin influences the proliferation of apical progenitors by amplifying Notch signaling (Lakomá et al., 2011). Intriguingly, Notch2 and Rbpj were amongst the strongly and significantly upregulated genes after circSlc45a4 knockdown in developing mouse cortex. The amplification of Notch signaling through reelin is promoting the symmetric proliferative divisions of apical progenitors (Lakomá et al., 2011). This could explain the observed increase in the number of apical progenitors, while there are less basal progenitors.

The here identified phenotype, caused by downregulation of circSLC45A4, is the first developmental brain phenotype described for a circRNA. However, the mechanism for this molecule remains elusive. Several potential modes of action have been investigated: interaction and regulation of miRNAs, interaction with RBPs, and translation (Figure 38 - B). Chimeric data obtained from mouse brain predicted an interaction of circSLC45A4 with miR-9, miR-135 and miR-181. miR-135 was previously shown to be involved in axon outgrowth (van Battum et al., 2018). Knockdown of circSLC45A4 in SH-SY5Y cells caused downregulation of miR-135. However, the interaction of circSLC45A4 with miR-135 could not be validated by pulldown of circSLC45A4. Furthermore, the downregulation of miR-135 was not recapitulated in embryonic mouse cortex after knockdown of circSlc45a4 (data not shown). Hence, a function of circSLC45A4 through interaction with miR-135 was ruled out. The changed expression levels of miR-135 after knockdown of circSLC45A4 were most likely secondary effects: hippocampal neurons downregulate miR-135 early during differentiation (van Battum et al., 2018).

Pulldown of circSLC45A4 was further used to identify RBPs interacting with the circular RNA by mass spectrometry. circSLC45A4 was strongly enriched after pulldown with the specific biotinylated antisense probes, but so was mRNA SLC45A4. Only 1 % of the ACTB present in the input was pulled down, however, given the abundance of ACTB per cell (~2500 molecules/cell, (Kislauskis et al., 1997)) 1 % (~25 molecules) of these transcripts would still be 6 times more than if 100 % of circSLC45A4 are pulled down (~4 copies per cell). Hence, it is

very likely that many of the 961 enriched proteins are not specific binders of circSLC45A4 but are bound by contaminating RNA molecules, even if they are strongly depleted compared to the input. Although some of the interactions identified by mass spectrometry could be validated by reverse pulldown, *i.e.* FXR2 and DOCK7, the interaction is not abundant and probably occurring too seldom to cause the observed phenotype when disrupted. However, FXR2 is a known regulator of BMP signaling in adult neurogenesis (Guo et al., 2011). Lowered activity levels of FXR2 causes an increase in Noggin expression, which in turn reduces BMP signaling and by that increases proliferation of neural stem/progenitor cells in the dentate gyrus. Indeed, an increase in Noggin expression was observed after cirSLC45A4 KD in mouse cortex. This could be causal for the observed increased neurogenesis or depletion of neural progenitors and should be further investigated.

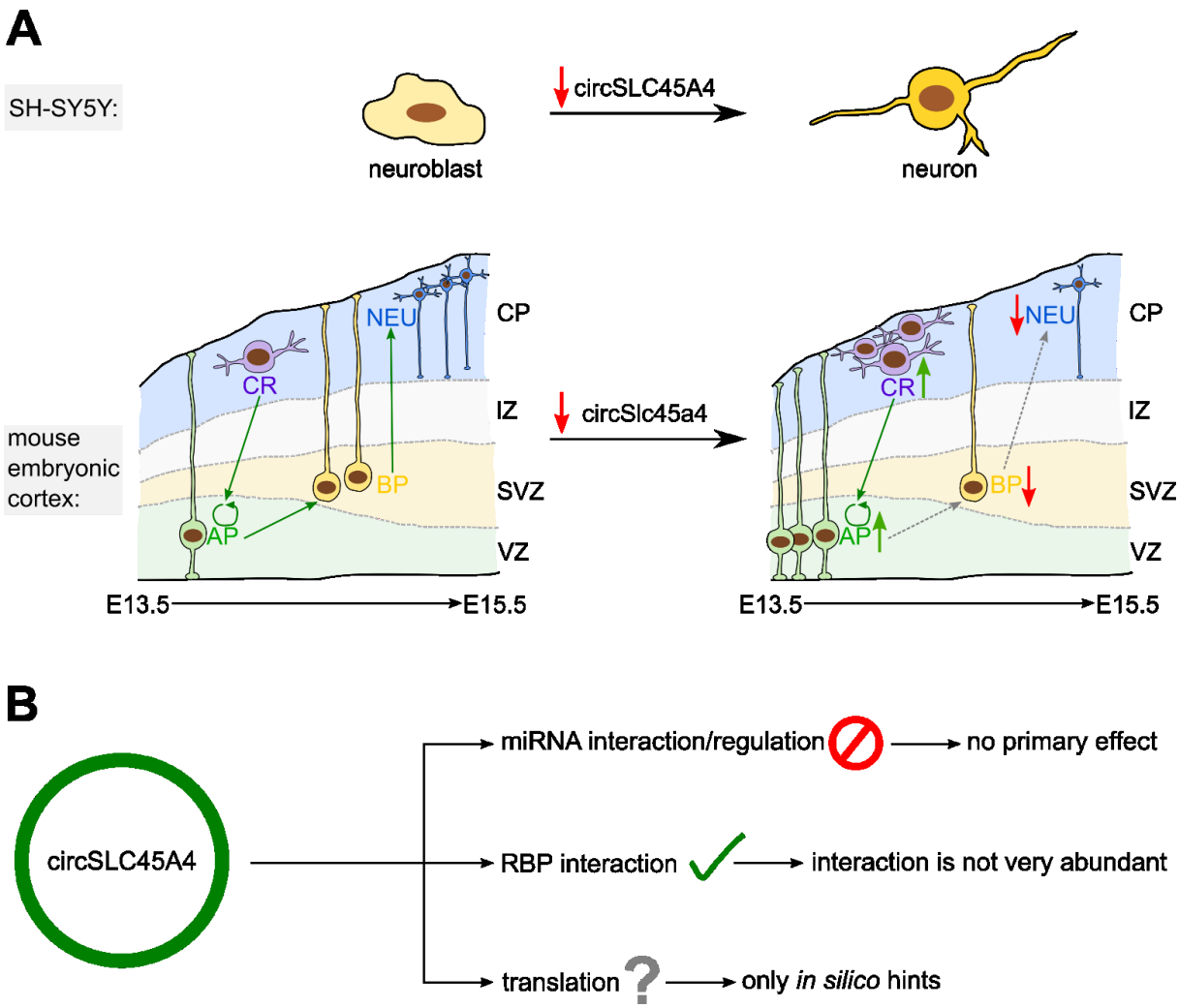


Figure 38: circSLC45A4 keeps neural cells in a progenitor state, although the mechanism remains elusive
(A) – Consequences of depleting circSLC45A4 in SH-SY5Y cells or in embryonic mouse cortex. **(B)** – Potential mechanisms of circSLC45A4.

Since there have been publications on the translation of circRNAs, the coding potential of circSLC45A4 was investigated as well. circSLC45A4 is produced from the first annotated exon of SLC45A4, which includes the 5'UTR and part of the CDS. Hence, the circRNA is sharing the canonical start codon with the mRNA. Surprisingly, the open reading frame would extend beyond the head-to-tail junction for 98 nt before running into 4 consecutive STOP codons. Conservation analysis showed that if this part beyond the head-to-tail junction was translated, it would be less conserved than the CDS of the mRNA on amino acid level. Furthermore, the chicken and frog (*Xenopus tropicalis*) open-reading frames would run into a STOP codon right after the head-to-tail junction. This is especially intriguing as a potential circSLC45A4 peptide would appear together with the six-layered mammalian cortex in evolution. However, attempts to validate the existence of a circSLC45A4 encoded peptide were not successful.

In summary, it was shown that circSLC45A4 depletion from either a human *in vitro* or a mouse *in vivo* differentiation system causes the depletion of the respective neural progenitor pools. This is the first developmental phenotype of a circular RNA identified. However, the mechanism remains elusive.

8.3 3D-seq: an innovative method to connect single-cell and spatial transcriptomics

The developing mouse cortex processed in this study is one example for a tissue where the spatial organization of cells and the knowledge thereof is of great importance. However, this can be generalized to virtually all multicellular tissues, as almost all of them are spatially organized which in turn influences the functionality of the tissue itself. For example, the stem cells of the intestinal crypt are localized at its basal tip and as differentiation of them progresses, they move to the apical tip of (Clevers, 2009). Yet, the knowledge about this spatial organization of the intestinal crypt derives from decades of research. Furthermore, researchers could take advantage of the fact that the tissue organization is highly reproducible in-between specimens of intestinal crypts. This is usually not the case for diseased tissues, especially tumors, which display considerable intra- and inter-tumor heterogeneity (Burrell et al., 2013) that complicates the research for effective cancer therapies. Moreover, a tumor cannot be considered as isolated malignant tissue but is influenced by its microenvironment and vice versa. Efforts have been made to use high resolution imaging of pathological sections to learn more about a tumor's microenvironment and combining this with independently obtained -omics methods. Nevertheless, features unique to a certain tumor specimen would be lost (Heindl et al., 2015).

Hence, there is a need for–omics methods that preserve spatial information, too. However, to date there is no method published that allows unbiased interrogation of any tissue with single-cell resolution while retrieving spatial information. 3D-seq offers a solution to this problem. This method is combining compartmentalization of a tissue slice with a combinatorial indexing protocol and therefore would enable unbiased measurement of the transcriptomes of single cells while their approximate spatial localization is measured on the same slice. Currently, 3D-seq allows the measurement of a few hundred genes per cell which is sufficient for cell type identification. Thus, for the first time 3D-seq allows the unbiased analysis of single-cell transcriptomes in their spatial context.

Furthermore, the 3D-seq protocol could be used in any lab with standard molecular biology equipment and access to an RNA-sequencing instrument. The only non-standard item used in this protocol is a 3D-printed plastic grid. FISH-based methods for example require highly specialized microscopy equipment (cf. 3 - Introduction: - 15 -3.3 - Investigating single-cell gene expression with spatial resolution). Also, 3D-seq is not limited to combining the spatial information with transcriptomic data. Other labeling strategies could be used to index proteins (CITE-seq based approach, (Stoeckius et al., 2017), or iTRAQ (Ross et al., 2004)) or DNA (tagmentation) in each compartment and therefore create tissue maps holding proteomic or genomic information.

It should be noted here, that the name "3D-seq" points to the intended future development of this method. Currently, the created tissue maps are two-dimensional. But the grid does not only provide a way to compartmentalize a single slice but could also be used as external coordinate system that is applied to several consecutive tissue slices. If these compartmentalized slices are then later on aligned *in silico* it would allow the three-dimensional reconstruction of a tissue with single-cell resolution.

Using the grid as external coordinate system for three-dimensional reconstruction is one advantage over simply cutting the tissue in small pieces. Instead of using one grid on consecutive tissue slices, one could also apply several grids with differing geometry to the same tissue slice consecutively in order to increase the resolution.

The grid is a central part of the 3D-seq protocol. Its design and production were limited to the methods available (ultimaker 2+ at BIMSB, Berlin; Projet HD 3000 at Fraunhofer, Potsdam). However, using a microfluidic chip as grid would have had the advantage of providing smaller grid compartments and that the loading of these grid compartments with the reaction solution could have been automated. Microfluidic chips are often manufactured from polydimethylsiloxane (PDMS) as it is a cheap, chemical inert, non-toxic, elastic, robust and thermostable material that allows building structures in the micrometer range. In addition, the surface of PDMS can be chemically modified to change the hydrophobic material's surface to become hydrophilic (Trantidou et al., 2017).

The 3D-seq protocol itself has been extensively modified starting from the published combinatorial indexing protocols (Cao et al., 2017; Rosenberg et al., 2018). Optimizations included the reverse transcription reaction, the ligation reaction, the bead purification after lysis, and the recovery of cells after centrifugation. Taken together, the modifications to the protocol resulted in a 195 % increase in material recovery compared to the methods used in the published combinatorial indexing protocols (Cao et al., 2017; Rosenberg et al., 2018).

However, the number of genes or UMIs/cell are still in the low hundreds. This number is of course strongly dependent on the number of reads spent per cell. In fact, the sequencing drives the costs of this protocol. While the library preparation for a single cell from a coronal adult mouse brain slices costs 2 cents/cell, sequencing the cells with 40.000 reads/cell on a Nextseq 500 adds 25 cents per cell. Downsampling showed, that doubling the number of reads per cell would increase the number of UMIs/cell by 10-20 % (Nikolaos Karaiskos, personal communication). Hence, spending more reads to improve the number of identified UMIs/cell would drive up the costs considerably. And, deeper sequencing and obtaining more UMIs/cell does not scale linearly, unfortunately. So, while deeper sequencing might provide additional information initially, this would soon saturate and therefore deeper sequencing is a limited solution to this problem and therefore not a viable option.

So, what is the reason for retrieving only a few hundred UMIs/cell with the current setup? Ideally, a UMI would be recovered for each transcript expressed within a cell. An estimated 50.000 – 300.000 transcripts are expressed in single GM12878 cells (B-lymphocytes) (Marinov et al., 2014). Each step of the protocol was thoroughly evaluated and optimized where possible. The ligation reaction is able to ligate the necessary adapter to 75 % of molecules. Hence, tagmentation or reverse transcription could be responsible for retrieving only a low percentage of actually expressed transcripts per cell. Tagmentation will approximately mark 50 % of the molecules in the reaction setup used for the 3D-seq protocol (Rykalina et al., 2017). Considering losing 25 % of transcripts due to ligation, 20 % with each bead purification (total of 5 in current protocol), and 50 % of transcripts due to tagmentation, this should still leave 12.3 % of the initial number of transcripts if RT were to work with 100 % efficiency. Hence, one could expect to recover 6150 UMIS/cell, if there were 50.000 transcripts expressed in a single

cell. However, only a mean of 158 UMIs/cell was recovered which would mean, that RT efficiency is only 2.6 % (Figure 39). This estimation of RT efficiency does not consider sequencing depth, which was at ~3500 reads/cell, or PCR overamplification to produce a sufficient amount of material to load a Nextseq 500. But even if PCR overamplification was 10-times, one would expect 350 UMIs/cell from 3500 reads/cell. Hence, RT efficiency could have been maximally 45.1 %. In conclusion, RT and tagmentation are the limiting steps for recovery of high numbers of UMIs/cell as they are the steps with the lowest efficiencies.

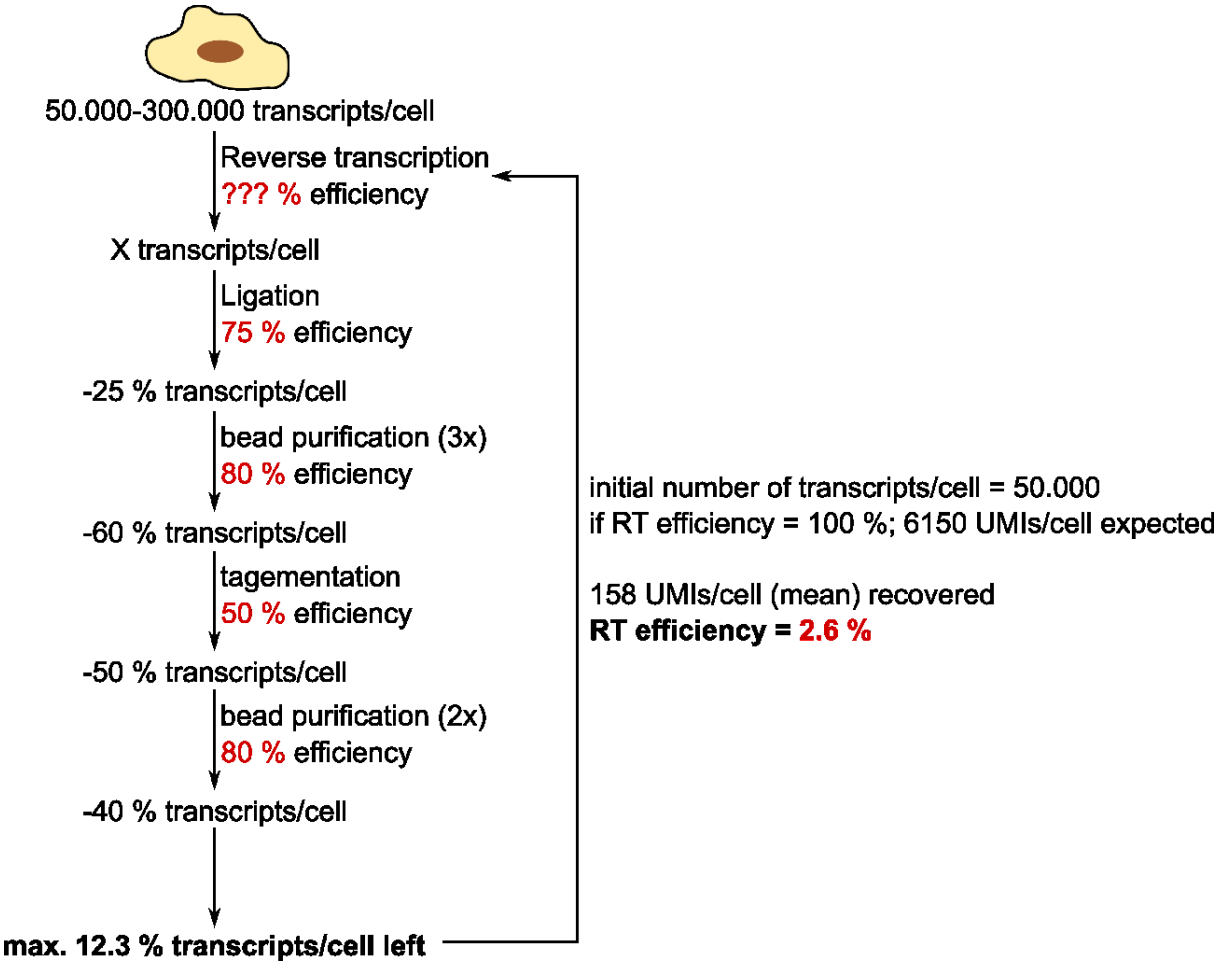


Figure 39: Estimation of RT efficiency in the 3D-seq protocol.
Material is lost at each enzymatic and purification step during the 3D-seq protocol to a different extent. Not considering the loss due to reverse transcription, a maximum of 12.3 % of the initial number of transcripts/cell would be left. With an initial 50.000 transcripts/cell one could expect up to 6150 UMIs/cell. However, only 158 UMIs/cell (mean) were recovered and would leave the RT efficiency at 2.6 %.

However, as the enzymes mediating these reactions will become more efficient in the future, so will the 3D-seq protocol. Furthermore, another reason for low efficiency of the *in situ* RT could be the limited penetrance of the cells after fixation and permeabilization. However, this step was already optimized.

But what can we already achieve with this kind of data? The aim of 3D-seq is the identification of spatial expression trends either on gene or on cell type level. 3D-seq is able to identify meaningful cell types even with a low number of UMIs/cell. Furthermore, spatial gene expression maps were created *in silico* based on the 3D-seq data. Zdhhc14, Myo16 and

Osbpl8 were found to be expressed in the expected compartments. However, one has to keep the low number of UMIs or genes/cell in mind as this will cause many cells to be false negatives. Hence, increasing the number of UMIs or genes/cell would yield more reliable and detailed gene expression maps. Additionally, it has to be stressed, that the shown maps are based on 6732 cells. However, in reality there will be approximately 50.000 - 75.000 cells within the grid. Estimating spatial trends from 10 % of the cells actually present in this tissue is limited in itself.

Taken together, 3D-seq is still the only method that would allow unbiased analysis of single-cell transcriptomes while preserving information on the original position of the cell within the tissue. The simplicity of 3D-seq is another advantage compared to other existing methods that try to address the same problem. 3D-seq only relies on the usage of an additional grid that can be tailored to the tissue's needs, has a short turnaround time (2 days from tissue fixation to RNA-sequencing), does not require any prior knowledge about the tissue of interest, e.g. which genes to stain by smFISH methods, is easily adaptable and versatile, and will allow the three dimensional reconstruction of a tissue of interest in the future.

8.4 Analyzing the neural transcriptional landscape in time and space

RNA-seq is a powerful tool to discover gene expression patterns in whole tissues or in single cells. Here it was applied to describe the versatile expression patterns of circular RNAs in a number of neural tissues, leading to the discovery of thousands of circular RNAs in human and mouse. In addition, many hints were collected that point at the complex regulation of circRNA expression and function in neural systems. Hence, a specific circular RNA, circSLC45A4, was studied in detail. This circRNA was shown to be an important regulator that maintains neural cells in their progenitor state, *in vitro* in human neuroblastoma cells, but also *in vivo* in the developing mouse cortex. To date, this is the only circular RNA described whose depletion causes a phenotype in neural development. But undoubtedly there will be additional examples of circular RNA functions in the future. Neural development, e.g. in the developing mouse cortex, is spatiotemporally organized. A method to investigate transcriptomes with single-cell and spatial resolution, 3D-seq, was developed. It was established on adult mouse brain but will have applications far beyond this well studied tissue. The method is especially suited for rare, highly variable samples of medical relevance, e.g. tumor samples, as it allows the retrieval of a maximum of information without the need for prior knowledge.

9 References

Bibliography

Abraham, M.A., and Lam, T.K.T. (2016). Glucagon action in the brain. Diabetologia *59*, 1367–1371.

ACDBio (2017). BaseScope™ Reagent Kit-RED | In Situ Hybridization, RNA-ISH | ACDBio.

Achim, K., Pettit, J.-B., Saraiva, L.R., Gavriouchkina, D., Larsson, T., Arendt, D., and Marioni, J.C. (2015). High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. Nat Biotechnol *33*, 503–509.

Adiconis, X., Borges-Rivera, D., Satija, R., DeLuca, D.S., Busby, M.A., Berlin, A.M., Sivachenko, A., Thompson, D.A., Wysoker, A., Fennell, T., et al. (2013). Comparative analysis of RNA sequencing methods for degraded or low-input samples. Nat Methods *10*, 623–629.

Aktaş, T., Avşar Ilık, İ., Maticzka, D., Bhardwaj, V., Pessoa Rodrigues, C., Mittler, G., Manke, T., Backofen, R., and Akhtar, A. (2017). DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. Nature *544*, 115–119.

Allen Institute for Brain Science (2011). Allen Mouse Brain Reference Atlas.

Alles, J., Karaiskos, N., Praktiknjo, S.D., Grosswendt, S., Wahle, P., Ruffault, P.-L., Ayoub, S., Schreyer, L., Boltengagen, A., Birchmeier, C., et al. (2017). Cell fixation and preservation for droplet-based single-cell transcriptomics. BMC Biol *15*, 44.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403–410.

Ameres, S.L., Horwich, M.D., Hung, J.-H., Xu, J., Ghildiyal, M., Weng, Z., and Zamore, P.D. (2010). Target RNA-directed trimming and tailing of small silencing RNAs. Science *328*, 1534–1539.

Amir, E.D., Davis, K.L., Tadmor, M.D., Simonds, E.F., Levine, J.H., Bendall, S.C., Shenfeld, D.K., Krishnaswamy, S., Nolan, G.P., and Pe'er, D. (2013). viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nat Biotechnol *31*, 545–552.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq — a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Angevine, J.B., and Sidman, R.L. (1961). Autoradiographic study of cell migration during histogenesis of cerebral cortex in the mouse. Nature *192*, 766–768.

Aprea, J., and Calegari, F. (2015). Long non-coding RNAs in corticogenesis: deciphering the non-coding code of the brain. EMBO J *34*, 2865–2884.

Aprea, J., Prenninger, S., Dori, M., Ghosh, T., Monasor, L.S., Wessendorf, E., Zocher, S., Massalini, S., Alexopoulou, D., Lesche, M., et al. (2013). Transcriptome sequencing during

mouse brain development identifies long non-coding RNAs functionally involved in neurogenic commitment. EMBO J *32*, 3145–3160.

Ashwal-Fluss, R., Meyer, M., Pamudurti, N.R., Ivanov, A., Bartok, O., Hanan, M., Evantal, N., Memczak, S., Rajewsky, N., and Kadener, S. (2014). circRNA biogenesis competes with pre-mRNA splicing. Mol Cell *56*, 55–66.

Baron, M., Veres, A., Wolock, S.L., Faust, A.L., Gaujoux, R., Vetere, A., Ryu, J.H., Wagner, B.K., Shen-Orr, S.S., Klein, A.M., et al. (2016). A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. Cell Syst. *3*, 346–360.e4.

Bartel, D.P. (2018). Metazoan MicroRNAs. Cell *173*, 20–51.

Van Battum, E.Y., Verhagen, M.G., Vangoor, V.R., Fujita, Y., Derijck, A.A.H.A., O'Duibhir, E., Giuliani, G., de Gunst, T., Adolfs, Y., Lelieveld, D., et al. (2018). An Image-Based miRNA Screen Identifies miRNA-135s As Regulators of CNS Axon Growth and Regeneration by Targeting Krüppel-like Factor 4. J Neurosci *38*, 613–630.

Bengtsson, M., Hemberg, M., Rorsman, P., and Ståhlberg, A. (2008). Quantification of mRNA in single cells and modelling of RT-qPCR induced noise. BMC Mol Biol *9*, 63.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing on JSTOR. J R Stat Soc Series B Stat Methodol *57*, 289–300.

Berget, S.M., Moore, C., and Sharp, P.A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci U S A *74*, 3171–3175.

Van den Berghe, V., Stappers, E., Vandesande, B., Dimidschstein, J., Kroes, R., Francis, A., Conidi, A., Lesage, F., Dries, R., Cazzola, S., et al. (2013). Directed migration of cortical interneurons depends on the cell-autonomous action of Sip1. Neuron *77*, 70–82.

BioNumbers BioNumbers - The Database of Useful Biological Numbers.

Boudreau, R.L., Spengler, R.M., Hylock, R.H., Kusenda, B.J., Davis, H.A., Eichmann, D.A., and Davidson, B.L. (2013). siSPOTR: a tool for designing highly specific and potent siRNAs for human and mouse. Nucleic Acids Res *41*, e9.

Braun, S., Domdey, H., and Wiebauer, K. (1996). Inverse splicing of a discontinuous pre-mRNA intron generates a circular exon in a HeLa cell nuclear extract. Nucleic Acids Res *24*, 4152–4157.

Bray, S.J. (2016). Notch signalling in context. Nat Rev Mol Cell Biol *17*, 722–735.

Briscoe, S.D., and Ragsdale, C.W. (2018). Homology, neocortex, and the evolution of developmental mechanisms. Science *362*, 190–193.

Britten, R.J., and Davidson, E.H. (1969). Gene regulation for higher cells: a theory. Science *165*, 349–357.

Brugiolo, M., Herzel, L., and Neugebauer, K.M. (2013). Counting on co-transcriptional splicing. F1000Prime Rep *5*, 9.

Burrell, R.A., McGranahan, N., Bartek, J., and Swanton, C. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. Nature *501*, 338–345.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol *36*, 411–420.

Cajal, R. y (1890). Sobre la existencia de células nerviosas especiales en la primera capa de las circumvoluciones cerebrales. Gaceta Medica Catalana *13*.

Cao, J., Packer, J.S., Ramani, V., Cusanovich, D.A., Huynh, C., Daza, R., Qiu, X., Lee, C., Furlan, S.N., Steemers, F.J., et al. (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. Science *357*, 661–667.

Capel, B., Swain, A., Nicolis, S., Hacker, A., Walter, M., Koopman, P., Goodfellow, P., and Lovell-Badge, R. (1993). Circular transcripts of the testis-determining gene Sry in adult mouse testis. Cell *73*, 1019–1030.

Chen, C.Y., and Sarnow, P. (1995). Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. Science *268*, 415–417.

Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015). RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science *348*, aaa6090.

Chen, Y.G., Kim, M.V., Chen, X., Batista, P.J., Aoyama, S., Wilusz, J.E., Iwasaki, A., and Chang, H.Y. (2017). Sensing self and foreign circular rnas by intron identity. Mol Cell *67*, 228–238.e5.

Chow, L.T., Gelinas, R.E., Broker, T.R., and Roberts, R.J. (1977). An amazing sequence arrangement at the 5’ ends of adenovirus 2 messenger RNA. Cell *12*, 1–8.

Clevers, H. (2009). Searching for adult stem cells in the intestine. EMBO Mol Med *1*, 255–259.

Clovis, Y.M., Enard, W., Marinaro, F., Huttner, W.B., and De Pietri Tonelli, D. (2012). Convergent repression of Foxp2 3’UTR by miR-9 and miR-132 in embryonic mouse neocortex: implications for radial migration of neurons. Development *139*, 3332–3342.

Cocquerelle, C., Daubersies, P., Majérus, M.A., Kerckaert, J.P., and Bailleul, B. (1992). Splicing with inverted order of exons occurs proximal to large introns. EMBO J *11*, 1095–1098.

Cocquerelle, C., Mascrez, B., Hétuin, D., and Bailleul, B. (1993). Mis-splicing yields circular RNA molecules. FASEB J *7*, 155–160.

Coggins, P.J., and Zwiers, H. (1991). B-50 (GAP-43): biochemistry and functional neurochemistry of a neuron-specific phosphoprotein. J Neurochem *56*, 1095–1106.

Conn, S.J., Pillman, K.A., Toubia, J., Conn, V.M., Salmanidis, M., Phillips, C.A., Roslan, S., Schreiber, A.W., Gregory, P.A., and Goodall, G.J. (2015). The RNA binding protein quaking regulates formation of circRNAs. Cell *160*, 1125–1134.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol *26*, 1367–1372.

Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics *13*, 2513–2526.

Crick, F. (1970). Central dogma of molecular biology. Nature *227*, 561–563.

Crick, F.H. (1958). On protein synthesis. Symp Soc Exp Biol *12*, 138–163.

Dahm, R., and Kiebler, M. (2005). Cell biology: silenced RNA on the move. Nature *438*, 432–435.

Danan, M., Schwartz, S., Edelheit, S., and Sorek, R. (2012). Transcriptome-wide discovery of circular RNAs in Archaea. Nucleic Acids Res *40*, 3131–3142.

Daniel, C., Silberberg, G., Behm, M., and Öhman, M. (2014). Alu elements shape the primate transcriptome by cis-regulation of RNA editing. Genome Biol *15*, R28.

Deininger, P. (2011). Alu elements: know the SINEs. Genome Biol *12*, 236.

Diener, T.O. (1989). Circular RNAs: relics of precellular evolution? Proc Natl Acad Sci U S A *86*, 9370–9374.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dotti, C.G., Sullivan, C.A., and Banker, G.A. (1988). The establishment of polarity by hippocampal neurons in culture. J Neurosci *8*, 1454–1468.

Doyle, M., and Kiebler, M.A. (2011). Mechanisms of dendritic mRNA transport and its role in synaptic tagging. EMBO J *30*, 3540–3552.

Du, W.W., Yang, W., Liu, E., Yang, Z., Dhaliwal, P., and Yang, B.B. (2016). Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2. Nucleic Acids Res *44*, 2846–2858.

Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M., and Coleman, P. (1992). Analysis of gene expression in single live neurons. Proc Natl Acad Sci U S A *89*, 3010–3014.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001a). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature *411*, 494–498.

Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. (2001b). Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. EMBO J *20*, 6877–6888.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Englund, C., Fink, A., Lau, C., Pham, D., Daza, R.A.M., Bulfone, A., Kowalczyk, T., and Hevner, R.F. (2005). Pax6, Tbr2, and Tbr1 are expressed sequentially by radial glia, intermediate progenitor cells, and postmitotic neurons in developing neocortex. J Neurosci *25*, 247–251.

Engreitz, J., Lander, E.S., and Guttman, M. (2015). RNA antisense purification (RAP) for mapping RNA interactions with chromatin. Methods Mol Biol *1262*, 183–197.

Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. (2003). MicroRNA targets in Drosophila. Genome Biol *5*, R1.

Enuka, Y., Lauriola, M., Feldman, M.E., Sas-Chen, A., Ulitsky, I., and Yarden, Y. (2016). Circular RNAs are long-lived and display only minimal early alterations in response to a growth factor. Nucleic Acids Res *44*, 1370–1383.

Fabian, M.R., Sonenberg, N., and Filipowicz, W. (2010). Regulation of mRNA translation and stability by microRNAs. Annu Rev Biochem *79*, 351–379.

Femino, A.M., Fay, F.S., Fogarty, K., and Singer, R.H. (1998). Visualization of single RNA transcripts in situ. Science *280*, 585–590.

Fietz, S.A., Kelava, I., Vogt, J., Wilsch-Bräuninger, M., Stenzel, D., Fish, J.L., Corbeil, D., Riehn, A., Distler, W., Nitsch, R., et al. (2010). OSVZ progenitors of human and ferret neocortex are epithelial-like and expand by integrin signaling. Nat Neurosci *13*, 690–699.

Finlay, B.L., and Darlington, R.B. (1995). Linked regularities in the development and evolution of mammalian brains. Science *268*, 1578–1584.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806–811.

Fischer, A.H., Jacobson, K.A., Rose, J., and Zeller, R. (2008). Cryosectioning Tissues. Cold Spring Harb Protoc *2008*, pdb.prot4991–pdb.prot4991.

Fukuchi-Shimogori, T., and Grove, E.A. (2001). Neocortex patterning by the secreted signaling molecule FGF8. Science *294*, 1071–1074.

Gautam, P., Anstey, K.J., Wen, W., Sachdev, P.S., and Cherbuin, N. (2015). Cortical gyrification and its relationships with cortical volume, cortical thickness, and cognitive performance in healthy mid-life adults. Behav Brain Res *287*, 331–339.

Götz, M., and Huttner, W.B. (2005). The cell biology of neurogenesis. Nat Rev Mol Cell Biol *6*, 777–788.

Gray, E.G., and Whittaker, V.P. (1962). The isolation of nerve endings from brain: an electron-microscopic study of cell fragments derived by homogenization and centrifugation. J Anat *96*, 79–88.

Grosswendt, S., Filipchyk, A., Manzano, M., Klironomos, F., Schilling, M., Herzog, M., Gottwein, E., and Rajewsky, N. (2014). Unambiguous identification of miRNA:target site interactions by different types of ligation reactions. Mol Cell *54*, 1042–1054.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell *35*, 849–857.

Guillemot, F., Lo, L.C., Johnson, J.E., Auerbach, A., Anderson, D.J., and Joyner, A.L. (1993). Mammalian achaete-scute homolog 1 is required for the early development of olfactory and autonomic neurons. Cell *75*, 463–476.

Guo, W., Zhang, L., Christopher, D.M., Teng, Z.-Q., Fausett, S.R., Liu, C., George, O.L., Klingensmith, J., Jin, P., and Zhao, X. (2011). RNA-binding protein FXR2 regulates adult hippocampal neurogenesis by reducing Noggin expression. Neuron *70*, 924–938.

Gupta, A., Tsai, L.-H., and Wynshaw-Boris, A. (2002). Life is a journey: a genetic look at neocortical development. Nat Rev Genet *3*, 342–355.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223–227.

Hamberger, A., Eriksson, O., and Norrby, K. (1971). Cell size distribution in brain suspensions and in fractions enriched with neuronal and glial cells. Exp Cell Res *67*, 380–388.

Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F., et al. (2018). Mapping the Mouse Cell Atlas by Microwell-Seq. Cell *172*, 1091–1107.e17.

Hansen, D.V., Lui, J.H., Parker, P.R.L., and Kriegstein, A.R. (2010). Neurogenic radial glia in the outer subventricular zone of human neocortex. Nature *464*, 554–561.

Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K., and Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. Nature *495*, 384–388.

Hatakeyama, J., Bessho, Y., Katoh, K., Ookawara, S., Fujioka, M., Guillemot, F., and Kageyama, R. (2004). Hes genes regulate size, shape and histogenesis of the nervous system by control of the timing of neural stem cell differentiation. Development *131*, 5539–5550.

Heindl, A., Nawaz, S., and Yuan, Y. (2015). Mapping spatial heterogeneity in the tumor microenvironment: a new era for digital pathology. Lab Invest *95*, 377–384.

Hentze, M.W., and Preiss, T. (2013). Circular RNAs: splicing's enigma variations. EMBO J *32*, 923–925.

Herculano-Houzel, S. (2009). The human brain in numbers: a linearly scaled-up primate brain. Front Hum Neurosci *3*, 31.

Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F., et al. (2006). The UCSC Genome Browser Database: update 2006. Nucleic Acids Res *34*, D590–8.

Le Hir, H., Saulière, J., and Wang, Z. (2016). The exon junction complex as a node of post-transcriptional networks. Nat Rev Mol Cell Biol *17*, 41–54.

Holt, C.E., and Bullock, S.L. (2009). Subcellular mRNA localization in animal cells and why it matters. Science *326*, 1212–1216.

Houseley, J.M., Garcia-Casado, Z., Pascual, M., Paricio, N., O'Dell, K.M.C., Monckton, D.G., and Artero, R.D. (2006). Noncanonical RNAs from transcripts of the Drosophila muscleblind gene. J Hered *97*, 253–260.

Hsu, M.T., and Coca-Prados, M. (1979). Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells. Nature *280*, 339–340.

Huang, C., Liang, D., Tatomer, D.C., and Wilusz, J.E. (2018). A length-dependent evolutionarily conserved pathway controls nuclear export of circular RNAs. Genes Dev *32*, 639–644.

Huang, H., Goto, M., Tsunoda, H., Sun, L., Taniguchi, K., Matsunaga, H., and Kambara, H. (2014). Non-biased and efficient global amplification of a single-cell cDNA library. Nucleic Acids Res *42*, e12.

Huttner, W.B., Schiebler, W., Greengard, P., and De Camilli, P. (1983). Synapsin I (protein I), a nerve terminal-specific phosphoprotein. III. Its association with synaptic vesicles studied in a highly purified synaptic vesicle preparation. J Cell Biol *96*, 1374–1388.

Ishihama, Y., Rappsilber, J., Andersen, J.S., and Mann, M. (2002). Microcolumns with self-assembled particle frits for proteomics. J Chromatogr A *979*, 233–239.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. Nat Methods *11*, 163–166.

Itoh, Y., Moriyama, Y., Hasegawa, T., Endo, T.A., Toyoda, T., and Gotoh, Y. (2013). Scratch regulates neuronal migration onset via an epithelial-mesenchymal transition-like mechanism. Nat Neurosci *16*, 416–425.

Ivanov, A., Memczak, S., Wyler, E., Torti, F., Porath, H.T., Orejuela, M.R., Piechotta, M., Levanon, E.Y., Landthaler, M., Dieterich, C., et al. (2015). Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. Cell Rep *10*, 170–177.

Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., et al. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science *343*, 776–779.

Jeck, W.R., Sorrentino, J.A., Wang, K., Slevin, M.K., Burd, C.E., Liu, J., Marzluff, W.F., and Sharpless, N.E. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. RNA *19*, 141–157.

Junker, J.P., Noël, E.S., Guryev, V., Peterson, K.A., Shah, G., Huisken, J., McMahon, A.P., Berezikov, E., Bakkers, J., and van Oudenaarden, A. (2014). Genome-wide RNA Tomography in the zebrafish embryo. Cell *159*, 662–675.

Kaech, S., and Banker, G. (2006). Culturing hippocampal neurons. Nat Protoc *1*, 2406–2415.

Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., Kocks, C., Rajewsky, N., and Zinzen, R.P. (2017). The *Drosophila* embryo at single-cell transcriptome resolution. Science *358*, 194–199.

Ke, R., Mignardi, M., Pacureanu, A., Svedlund, J., Botling, J., Wählby, C., and Nilsson, M. (2013). In situ sequencing for RNA analysis in preserved tissue and cells. Nat Methods *10*, 857–860.

Kirischuk, S., Luhmann, H.J., and Kilb, W. (2014). Cajal-Retzius cells: update on structural and functional properties of these mystic neurons that bridged the 20th century. Neuroscience *275*, 33–46.

Kislauskis, E.H., Zhu, X., and Singer, R.H. (1997). beta-Actin messenger RNA localization and protein synthesis augment cell motility. J Cell Biol *136*, 1263–1270.

Kivioja, T., Vähärautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S., and Taipale, J. (2011). Counting absolute numbers of molecules using unique molecular identifiers. Nat Methods *9*, 72–74.

Kleaveland, B., Shi, C.Y., Stefano, J., and Bartel, D.P. (2018). A network of noncoding regulatory rnas acts in the mammalian brain. Cell *174*, 350–362.e17.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell *161*, 1187–1201.

Komuro, H., and Rakic, P. (1998). Distinct modes of neuronal migration in different domains of developing cerebellar cortex. J Neurosci *18*, 1478–1490.

Konarska, M., Filipowicz, W., Domdey, H., and Gross, H.J. (1981). Binding of ribosomes to linear and circular forms of the 5'-terminal leader fragment of tobacco-mosaic-virus RNA. Eur J Biochem *114*, 221–227.

Kos, A., Dijkema, R., Arnberg, A.C., van der Meide, P.H., and Schellekens, H. (1986). The hepatitis delta (delta) virus possesses a circular RNA. Nature *323*, 558–560.

Kovalevich, J., and Langford, D. (2013). Considerations for the use of SH-SY5Y neuroblastoma cells in neurobiology. Methods Mol Biol *1078*, 9–21.

Kozak, M. (1979). Inability of circular mRNA to attach to eukaryotic ribosomes. Nature *280*, 82–85.

Krek, A., Grün, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M., et al. (2005). Combinatorial microRNA target predictions. Nat Genet *37*, 495–500.

Kriegstein, A.R., and Noctor, S.C. (2004). Patterns of neuronal migration in the embryonic cortex. Trends Neurosci *27*, 392–399.

Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. Cell *31*, 147–157.

Krüger, J., and Rehmsmeier, M. (2006). RNAhybrid: microRNA target prediction easy, fast and flexible. Nucleic Acids Res *34*, W451–4.

Kwan, K.Y., Sestan, N., and Anton, E.S. (2012). Transcriptional co-regulation of neuronal migration and laminar identity in the neocortex. Development *139*, 1535–1546.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680–685.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853–858.

Lakomá, J., Garcia-Alonso, L., and Luque, J.M. (2011). Reelin sets the pace of neocortical neurogenesis. Development *138*, 5223–5234.

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. PLoS Comput Biol *9*, e1003118.

Lee, V.M., and Andrews, P.W. (1986). Differentiation of NTERA-2 clonal human embryonal carcinoma cells into neurons involves the induction of all three neurofilament proteins. J Neurosci *6*, 514–521.

Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Yang, J.L., Ferrante, T.C., Terry, R., Jeanty, S.S.F., Li, C., Amamoto, R., et al. (2014). Highly multiplexed subcellular RNA sequencing in situ. Science *343*, 1360–1363.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843–854.

Legnini, I., Di Timoteo, G., Rossi, F., Morlando, M., Briganti, F., Sthandier, O., Fatica, A., Santini, T., Andronache, A., Wade, M., et al. (2017). Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. Mol Cell *66*, 22–37.e9.

Lein, E., Borm, L.E., and Linnarsson, S. (2017). The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. Science *358*, 64–69.

Levanon, E.Y., Eisenberg, E., Yelin, R., Nemzer, S., Hallegger, M., Shemesh, R., Fligelman, Z.Y., Shoshan, A., Pollock, S.R., Sztybel, D., et al. (2004). Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat Biotechnol *22*, 1001–1005.

Levesque, M.J., Ginart, P., Wei, Y., and Raj, A. (2013). Visualizing SNVs to quantify allele-specific expression in single cells. Nat Methods *10*, 865–867.

Li, X., Liu, C.-X., Xue, W., Zhang, Y., Jiang, S., Yin, Q.-F., Wei, J., Yao, R.-W., Yang, L., and Chen, L.-L. (2017). Coordinated circRNA Biogenesis and Function with NF90/NF110 in Viral Infection. Mol Cell *67*, 214–227.e7.

Liang, D., and Wilusz, J.E. (2014). Short intronic repeat sequences facilitate circular RNA production. Genes Dev *28*, 2233–2247.

Liang, D., Tatomer, D.C., Luo, Z., Wu, H., Yang, L., Chen, L.-L., Cherry, S., and Wilusz, J.E. (2017). The Output of Protein-Coding Genes Shifts to Circular RNAs When the Pre-mRNA Processing Machinery Is Limiting. Mol Cell *68*, 940–954.e3.

Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res *41*, e108.

Linnarson lab mousebrain.org/.

Love, M.I., Huber, W., and Anders, S. (2014a). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550–550.

Love, M.I., Huber, W., and Anders, S. (2014b). Moderated estimation of fold change and' ' dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Lubeck, E., and Cai, L. (2012). Single-cell systems biology by super-resolution imaging and combinatorial labeling. Nat Methods *9*, 743–748.

Lubeck, E., Coskun, A.F., Zhiyentayev, T., Ahmad, M., and Cai, L. (2014). Single-cell in situ RNA profiling by sequential hybridization. Nat Methods *11*, 360–361.

Van der Maaten, L., and Hinton, G. (2008). Visualizing Data using t-SNE. J Mach Learn Res *9*, 2579–2605.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214.

Marinov, G.K., Williams, B.A., McCue, K., Schroth, G.P., Gertz, J., Myers, R.M., and Wold, B.J. (2014). From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. Genome Res *24*, 496–510.

Dal Maschio, M., Ghezzi, D., Bony, G., Alabastri, A., Deidda, G., Brondi, M., Sato, S.S., Zaccaria, R.P., Di Fabrizio, E., Ratto, G.M., et al. (2012). High-performance and site-directed in utero electroporation by a triple-electrode probe. Nat Commun *3*, 960.

De la Mata, M., Gaidatzis, D., Vitanescu, M., Stadler, M.B., Wentzel, C., Scheiffele, P., Filipowicz, W., and Großhans, H. (2015). Potent degradation of neuronal miRNAs induced by highly complementary targets. EMBO Rep *16*, 500–511.

Meister, G., Landthaler, M., Dorsett, Y., and Tuschl, T. (2004). Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing. RNA *10*, 544–550.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. Nature *495*, 333–338.

Menezes, J.R., and Luskin, M.B. (1994). Expression of neuron-specific tubulin defines a novel population in the proliferative layers of the developing telencephalon. J Neurosci *14*, 5399–5416.

Mercer, T.R., Dinger, M.E., and Mattick, J.S. (2009). Long non-coding RNAs: insights into functions. Nat Rev Genet *10*, 155–159.

Merrick, W.C. (2004). Cap-dependent and cap-independent translation in eukaryotic systems. Gene *332*, 1–11.

Meyer, K.D., Patil, D.P., Zhou, J., Zinoviev, A., Skabkin, M.A., Elemento, O., Pestova, T.V., Qian, S.-B., and Jaffrey, S.R. (2015). 5' UTR m(6)A Promotes Cap-Independent Translation. Cell *163*, 999–1010.

Mi, D., Li, Z., Lim, L., Li, M., Moissidis, M., Yang, Y., Gao, T., Hu, T.X., Pratt, T., Price, D.J., et al. (2018). Early emergence of cortical interneuron diversity in the mouse embryo. Science *360*, 81–85.

Mihalas, A.B., Elsen, G.E., Bedogni, F., Daza, R.A.M., Ramos-Laguna, K.A., Arnold, S.J., and Hevner, R.F. (2016). Intermediate progenitor cohorts differentially generate cortical layers and require tbr2 for timely acquisition of neuronal subtype identity. Cell Rep *16*, 92–105.

Molyneaux, B.J., Arlotta, P., Menezes, J.R.L., and Macklis, J.D. (2007). Neuronal subtype specification in the cerebral cortex. Nat Rev Neurosci *8*, 427–437.

Moncada, R., Chiodin, M., Devlin, J.C., Baron, M., Hajdu, C.H., Simeone, D., and Yanai, I. (2018). Building a tumor atlas: integrating single-cell RNA-Seq data with spatial transcriptomics in pancreatic ductal adenocarcinoma. BioRxiv.

Moore, M.J., Scheel, T.K.H., Luna, J.M., Park, C.Y., Fak, J.J., Nishiuchi, E., Rice, C.M., and Darnell, R.B. (2015). miRNA-target chimeras reveal miRNA 3'-end pairing as a major determinant of Argonaute target specificity. Nat Commun *6*, 8864.

Mora-Bermúdez, F., García, M.T., and Huttner, W.B. (2013). Stem cells: neural stem cells in cerebral cortex development. In Neuroscience in the 21st Century, D.W. Pfaff, ed. (New York, NY: Springer New York), pp. 137–159.

Morris, K.V., and Mattick, J.S. (2014). The rise of regulatory RNA. Nat Rev Genet *15*, 423–437.

Muraro, M.J., Dharmadhikari, G., Grün, D., Groen, N., Dielen, T., Jansen, E., van Gurp, L., Engelse, M.A., Carlotti, F., de Koning, E.J.P., et al. (2016). A Single-Cell Transcriptome Atlas of the Human Pancreas. Cell Syst. *3*, 385–394.e3.

Nakayama, K.I., and Nakayama, K. (2006). Ubiquitin ligases: cell-cycle control and cancer. Nat Rev Cancer *6*, 369–381.

Nigro, J.M., Cho, K.R., Fearon, E.R., Kern, S.E., Ruppert, J.M., Oliner, J.D., Kinzler, K.W., and Vogelstein, B. (1991). Scrambled exons. Cell *64*, 607–613.

Noctor, S.C., Martinez-Cerdeño, V., and Kriegstein, A.R. (2007). Neural stem and progenitor cells in cortical development. Novartis Found Symp *288*, 59–73; discussion 73.

Northcutt, R.G. (2002). Understanding vertebrate brain evolution. Integr Comp Biol *42*, 743–756.

O'Brien, R.J., Xu, D., Petralia, R.S., Steward, O., Huganir, R.L., and Worley, P. (1999). Synaptic clustering of AMPA receptors by the extracellular immediate-early gene product Narp. Neuron *23*, 309–323.

Okuda, H., Xing, F., Pandey, P.R., Sharma, S., Watabe, M., Pai, S.K., Mo, Y.-Y., Iiizumi-Gairani, M., Hirota, S., Liu, Y., et al. (2013). miR-7 suppresses brain metastasis of breast cancer stem-like cells by modulating KLF4. Cancer Res *73*, 1434–1444.

Osenberg, S., Paz Yaacov, N., Safran, M., Moshkovitz, S., Shtrichman, R., Sherf, O., Jacob-Hirsch, J., Keshet, G., Amariglio, N., Itskovitz-Eldor, J., et al. (2010). Alu sequences in undifferentiated human embryonic stem cells display high levels of A-to-I RNA editing. PLoS ONE *5*, e11173.

Påhlman, S., Ruusala, A.I., Abrahamsson, L., Mattsson, M.E., and Esscher, T. (1984). Retinoic acid-induced differentiation of cultured human neuroblastoma cells: a comparison with phorbolester-induced differentiation. Cell Differ *14*, 135–144.

Pamudurti, N.R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., Hanan, M., Wyler, E., Perez-Hernandez, D., Ramberger, E., et al. (2017). Translation of CircRNAs. Mol Cell *66*, 9–21.e7.

Paridaen, J.T.M.L., and Huttner, W.B. (2014). Neurogenesis during development of the vertebrate central nervous system. EMBO Rep *15*, 351–364.

Pasman, Z., Been, M.D., and Garcia-Blanco, M.A. (1996). Exon circularization in mammalian nuclear extracts. RNA *2*, 603–610.

Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science *344*, 1396–1401.

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res *29*, e45.

Piwecka, M., Glažar, P., Hernandez-Miranda, L.R., Memczak, S., Wolf, S.A., Rybak-Wolf, A., Filipchyk, A., Klironomos, F., Cerda Jara, C.A., Fenske, P., et al. (2017). Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function. Science *357*.

Pleasure, S.J., and Lee, V.M. (1993). NTera 2 cells: a human cell line which displays characteristics expected of a human committed neuronal progenitor cell. J Neurosci Res *35*, 585–602.

Podrygajlo, G., Tegenge, M.A., Gierse, A., Paquet-Durand, F., Tan, S., Bicker, G., and Stern, M. (2009). Cellular phenotypes of human model neurons (NT2) after differentiation in aggregate culture. Cell Tissue Res *336*, 439–452.

Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res *20*, 110–121.

Polleux, F., Whitford, K.L., Dijkhuizen, P.A., Vitalis, T., and Ghosh, A. (2002). Control of cortical interneuron migration by neurotrophins and PI3-kinase signaling. Development *129*, 3147–3160.

Poulin, J.-F., Tasic, B., Hjerling-Leffler, J., Trimarchi, J.M., and Awatramani, R. (2016). Disentangling neural cell diversity using single-cell transcriptomics. Nat Neurosci *19*, 1131–1141.

Purves, D., Augustine, G.J., Fitzpatrick, D., Hall, W.C., LaMantia, A.-S., McNamara, J.O., and Williams, S.M. (2004). Neuroscience (Sinauer Associates, Inc.).

Raj, A., and Tyagi, S. (2010). Detection of individual endogenous RNA transcripts in situ using multiple singly labeled probes. Meth Enzymol *472*, 365–386.

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods *5*, 877–879.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc *2*, 1896–1906.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature *403*, 901–906.

Retzius, G. (1893). Die Cajal'schen Zellen der Grosshirnrinde beim Menschen und bei Säiugerthieren. Biologisches Untersuchungen *5*.

Rieder, L.E., Savva, Y.A., Reyna, M.A., Chang, Y.-J., Dorsky, J.S., Rezaei, A., and Reenan, R.A. (2015). Dynamic response of RNA editing to temperature in Drosophila. BMC Biol *13*, 1.

Ringnér, M. (2008). What is principal component analysis? Nat Biotechnol *26*, 303–304.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res *43*, e47.

Rosenberg, A.B., Roco, C.M., Muscat, R.A., Kuchina, A., Sample, P., Yao, Z., Graybuck, L.T., Peeler, D.J., Mukherjee, S., Chen, W., et al. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science *360*, 176–182.

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., et al. (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics *3*, 1154–1169.

Ross, R.A., Spengler, B.A., and Biedler, J.L. (1983). Coordinate morphological and biochemical interconversion of human neuroblastoma cells. J Natl Cancer Inst *71*, 741–747.

Rüegger, S., and Großhans, H. (2012). MicroRNA turnover: when, how, and why. Trends Biochem Sci *37*, 436–446.

Rybak-Wolf, A., Stottmeister, C., Glažar, P., Jens, M., Pino, N., Giusti, S., Hanan, M., Behm, M., Bartok, O., Ashwal-Fluss, R., et al. (2015). Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. Mol Cell *58*, 870–885.

Rykalina, V., Shadrin, A., Lehrach, H., and Borodina, T. (2017). qPCR-based characterization of DNA fragmentation efficiency of Tn5 transposomes. Biol. Methods Protoc. *2*.

Sahara, S., and O'Leary, D.D.M. (2009). Fgf10 regulates transition period of cortical stem cell differentiation to radial glia controlling generation of neurons and basal progenitors. Neuron *63*, 48–62.

Saito, T., and Nakatsuji, N. (2001). Efficient gene transfer into the embryonic mouse brain using in vivo electroporation. Dev Biol *240*, 237–246.

Salomoni, P., and Calegari, F. (2010). Cell cycle control of mammalian neural stem cells: putting a speed limit on G1. Trends Cell Biol *20*, 233–243.

Salzman, J., Gawad, C., Wang, P.L., Lacayo, N., and Brown, P.O. (2012). Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. PLoS ONE *7*, e30733.

Salzman, J., Chen, R.E., Olsen, M.N., Wang, P.L., and Brown, P.O. (2013). Cell-type specific features of circular RNA expression. PLoS Genet *9*, e1003777.

Sanger, H.L., Klotz, G., Riesner, D., Gross, H.J., and Kleinschmidt, A.K. (1976). Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. Proc Natl Acad Sci U S A *73*, 3852–3856.

Sanz, E., Yang, L., Su, T., Morris, D.R., McKnight, G.S., and Amieux, P.S. (2009). Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. Proc Natl Acad Sci U S A *106*, 13939–13944.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat Biotechnol *33*, 495–502.

Savva, Y.A., Rieder, L.E., and Reenan, R.A. (2012). The ADAR protein family. Genome Biol *13*, 252.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat Methods *9*, 676–682.

Schmid, C.W. (1996). Alu: Structure, Origin, Evolution, Significance, and Function of One-Tenth of Human DNA. (Elsevier), pp. 283–319.

Schuman, E.M. (1999). mRNA trafficking and local protein synthesis at the synapse. Neuron *23*, 645–648.

Sobczak, M., Chumak, V., Pomorski, P., Wojtera, E., Majewski, Ł., Nowak, J., Yamauchi, J., and Rędowicz, M.J. (2016). Interaction of myosin VI and its binding partner DOCK7 plays an important role in NGF-stimulated protrusion formation in PC12 cells. Biochim Biophys Acta *1863*, 1589–1600.

Sokol, A.M., Sztolsztener, M.E., Wasilewski, M., Heinz, E., and Chacinska, A. (2014). Mitochondrial protein translocases for survival and wellbeing. FEBS Lett *588*, 2484–2495.

Stagni, F., Giacomini, A., Guidi, S., Ciani, E., and Bartesaghi, R. (2015). Timing of therapies for Down syndrome: the sooner, the better. Front Behav Neurosci *9*, 265.

Stagsted, L.V.W., Nielsen, K.M., Daugaard, I., and Hansen, T.B. (2018). Non-coding AUG circRNAs constitute an abundant and conserved subclass of circles. BioRxiv.

Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science *353*, 78–82.

Starke, S., Jost, I., Rossbach, O., Schneider, T., Schreiner, S., Hung, L.-H., and Bindereif, A. (2015). Exon circularization requires canonical splice signals. Cell Rep *10*, 103–111.

Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P.K., Swerdlow, H., Satija, R., and Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. Nat Methods *14*, 865–868.

Striedter, G.F. (2005). Principles of brain evolution (Sunderland, Mass: Sinauer Associates).

Sun, T., and Hevner, R.F. (2014). Growth and folding of the mammalian cerebral cortex: from molecules to malformations. Nat Rev Neurosci *15*, 217–232.

Suzuki, H., Zuo, Y., Wang, J., Zhang, M.Q., Malhotra, A., and Mayeda, A. (2006). Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. Nucleic Acids Res *34*, e63.

Tabata, H., and Nakajima, K. (2001). Efficient in utero gene transfer system to the developing mouse brain using electroporation: visualization of neuronal migration in the developing cortex. Neuroscience *103*, 865–872.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods *6*, 377–382.

Taniguchi, K., Kajiyama, T., and Kambara, H. (2009). Quantitative analysis of gene expression in a single cell by qPCR. Nat Methods *6*, 503–506.

Taverna, E., Götz, M., and Huttner, W.B. (2014). The cell biology of neurogenesis: toward an understanding of the development and evolution of the neocortex. Annu Rev Cell Dev Biol *30*, 465–502.

ThermoScientific (2017). Maxima H Minus Reverse Transcriptase.

Trantidou, T., Elani, Y., Parsons, E., and Ces, O. (2017). Hydrophilic surface modification of PDMS for droplet microfluidics using a simple, quick, and robust method via PVA deposition. Microsyst. Nanoeng. *3*, 16091.

Tsien, J.Z. (2016). Cre-Lox Neurogenetics: 20 Years of Versatile Applications in Brain Research and Counting…. Front Genet *7*, 19.

Tsui, C.C., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., Barnes, C., and Worley, P.F. (1996). Narp, a novel member of the pentraxin family, promotes neurite outgrowth and is dynamically regulated by neuronal activity. J Neurosci *16*, 2463–2478.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods *13*, 731–740.

Vargas, J., Herrera, L., and Mireles, J. (2014). WOHLERS REPORT Appendix E: Material properties.

Venø, M.T., Hansen, T.B., Venø, S.T., Clausen, B.H., Grebing, M., Finsen, B., Holm, I.E., and Kjems, J. (2015). Spatio-temporal regulation of circular RNA expression during porcine embryonic brain development. Genome Biol *16*, 245.

Vitak, S.A., Torkenczy, K.A., Rosenkrantz, J.L., Fields, A.J., Christiansen, L., Wong, M.H., Carbone, L., Steemers, F.J., and Adey, A. (2017). Sequencing thousands of single-cell genomes with combinatorial indexing. Nat Methods *14*, 302–308.

Vitavska, O., and Wieczorek, H. (2017). Putative role of an SLC45 H(+)/sugar cotransporter in mammalian spermatozoa. Pflugers Arch.

Wagner, G.P., Kin, K., and Lynch, V.J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci *131*, 281–285.

Wahlstedt, H., Daniel, C., Ensterö, M., and Ohman, M. (2009). Large-scale mRNA sequencing determines global regulation of RNA editing during brain development. Genome Res *19*, 978–986.

Wang, Y., and Wang, Z. (2015). Efficient backsplicing produces translatable circular mRNAs. RNA *21*, 172–179.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470–476.

Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., and He, C. (2015). N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. Cell *161*, 1388–1399.

Warner, J.R., Soeiro, R., Birnboim, H.C., Girard, M., and Darnell, J.E. (1966). Rapidly labeled HeLa cell nuclear RNA. I. Identification by zone sedimentation of a heterogeneous fraction separate from ribosomal precursor RNA. J Mol Biol *19*, 349–361.

Warren, L., Bryder, D., Weissman, I.L., and Quake, S.R. (2006). Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. Proc Natl Acad Sci U S A *103*, 17807–17812.

Westholm, J.O., Miura, P., Olson, S., Shenker, S., Joseph, B., Sanfilippo, P., Celniker, S.E., Graveley, B.R., and Lai, E.C. (2014). Genome-wide analysis of drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. Cell Rep *9*, 1966–1980.

Wilkins, M.R., Lindskog, I., Gasteiger, E., Bairoch, A., Sanchez, J.C., Hochstrasser, D.F., and Appel, R.D. (1997). Detailed peptide characterization using PEPTIDEMASS--a World-Wide-Web-accessible tool. Electrophoresis *18*, 403–408.

Wilkins, M.R., Gasteiger, E., Bairoch, A., Sanchez, J.C., Williams, K.L., Appel, R.D., and Hochstrasser, D.F. (1999). Protein identification and analysis tools in the ExPASy server. Methods Mol Biol *112*, 531–552.

Wonders, C.P., and Anderson, S.A. (2006). The origin and specification of cortical interneurons. Nat Rev Neurosci *7*, 687–696.

Worthington Biochemicals Worthington Tissue Dissociation Guide.

Xu, D., Hopf, C., Reddy, R., Cho, R.W., Guo, L., Lanahan, A., Petralia, R.S., Wenthold, R.J., O'Brien, R.J., and Worley, P. (2003). Narp and NP1 form heterocomplexes that function in developmental and activity-dependent synaptic plasticity. Neuron *39*, 513–528.

Xu, Y., Ong, J., Guan, S., and Nichols, N. (2017). US9580698B1 - Mutant reverse transcriptase - Google Patents (US).

Yang, Y., Fan, X., Mao, M., Song, X., Wu, P., Zhang, Y., Jin, Y., Yang, Y., Chen, L.-L., Wang, Y., et al. (2017). Extensive translation of circular RNAs driven by N6-methyladenosine. Cell Res *27*, 626–641.

Yeo, G., Holste, D., Kreiman, G., and Burge, C.B. (2004). Variation in alternative splicing across human tissues. Genome Biol *5*, R74.

Yokota, Y., Gashghaei, H.T., Han, C., Watson, H., Campbell, K.J., and Anton, E.S. (2007). Radial glial dependent and independent dynamics of interneuronal migration in the developing cerebral cortex. PLoS ONE *2*, e794.

You, X., Vlatkovic, I., Babic, A., Will, T., Epstein, I., Tushev, G., Akbalik, G., Wang, M., Glock, C., Quedenau, C., et al. (2015). Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. Nat Neurosci *18*, 603–610.

Yuzwa, S.A., Borrett, M.J., Innes, B.T., Voronova, A., Ketela, T., Kaplan, D.R., Bader, G.D., and Miller, F.D. (2017). Developmental Emergence of Adult Neural Stem Cells as Revealed by Single-Cell Transcriptional Profiling. Cell Rep *21*, 3970–3986.

Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138–1142.

Zhang, X., Huang, C.T., Chen, J., Pankratz, M.T., Xi, J., Li, J., Yang, Y., Lavaute, T.M., Li, X.-J., Ayala, M., et al. (2010). Pax6 is a human neuroectoderm cell fate determinant. Cell Stem Cell *7*, 90–100.

Zhang, X.-O., Wang, H.-B., Zhang, Y., Lu, X., Chen, L.-L., and Yang, L. (2014). Complementary sequence-mediated exon circularization. Cell *159*, 134–147.

Zhang, Y., Xue, W., Li, X., Zhang, J., Chen, S., Zhang, J.-L., Yang, L., and Chen, L.-L. (2016). The biogenesis of nascent circular rnas. Cell Rep *15*, 611–624.

Zhao, X.-D., Lu, Y.-Y., Guo, H., Xie, H.-H., He, L.-J., Shen, G.-F., Zhou, J.-F., Li, T., Hu, S.-J., Zhou, L., et al. (2015). MicroRNA-7/NF-κB signaling regulatory feedback circuit regulates gastric carcinogenesis. J Cell Biol *210*, 613–627.

Zhu, Y.Y., Machleder, E.M., Chenchik, A., Li, R., and Siebert, P.D. (2001). Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. BioTechniques *30*, 892–897.

10 Appendix

10.1 Oligonucleotides

Name	Sequence 5'→ 3'	Purpose
Stellaris circSLC_1	TCACAACGGTATGAGACATG	Stellaris smFISH staining of circSLC45A4
Stellaris circSLC_2	CAAGTGGCTATCTCACAACT	
Stellaris circSLC_3	AACACTTACATTCCTCTTTG	
Stellaris circSLC_4	CTGCTGTGTTCCTCGGGCAG	
Stellaris circSLC_5	AAGACGTCTTCTCTTTCTGC	
Stellaris circSLC_6	TCATATATCTGTAATGATAA	
Stellaris circSLC_7	TCTATCTATATAAATAATGC	
Stellaris circSLC_9	GGAGCCATTTTCATTACCAC	
Stellaris circSLC_10	AGATTCCGGGTCGGCATTCT	
Stellaris circSLC_11	CGGATAACTCTTGAACTTGC	
Stellaris circSLC_12	TTCTGCGGGTCCGGCAGGGG	
Stellaris circSLC_13	CTCGGCCTCTGCGCCTCCGG	
Stellaris circSLC_14	CCTCGCTGATGGTCTCGCAG	
Stellaris circSLC_15	ATGGGGATTCGGTCTATGGA	
Stellaris circSLC_16	CCCGTGCATCACCCACAGGC	
Stellaris circSLC_17	ACTCCCTGCCAAACATCACC	
Stellaris circSLC_18	GCGGTTTCCATGGCGTAACA	
Stellaris circSLC_19	CAACAGTATTGGTGTGACCA	
Stellaris circSLC_20	GAGTGACTGGATGATCAATC	
Stellaris circSLC_21	TGGTCCGCTGGTAATCCCCA	
Stellaris circSLC_22	GTGGGAAGGTGATACAAACA	
Stellaris circSLC_23	TGATTTTTCTTGCTACAGAG	
Stellaris circSLC_24	TATTACTGAAGAGGTGCTGA	
Stellaris circSLC_25	CAGTTAATACCAGTTTCATC	
Stellaris circSLC_26	TTAAAAATCCATAATTGCCA	
Stellaris circSLC_27	GCTGGTTTACTGTGAATTAG	
Stellaris circSLC_28	TTGGATTTTAAACATATGGC	
Stellaris circSLC_29	AGCAACACAGAACGATTTTT	
Stellaris mRNA SLC_1	GTGAAGATGAGGCCAAGGAT	Stellaris smFISH staining of mRNA SLC45A4
Stellaris mRNA SLC_2	ACTCGCAGACCCAATGAGAG	
Stellaris mRNA SLC_3	CAGAGGGCGAGGATGAAGGG	
Stellaris mRNA SLC_4	AACGCCAAAGAGGACGCCAA	
Stellaris mRNA SLC_5	GGCAGAGCCGTTAAGGAAAA	
Stellaris mRNA SLC_6	CGTGAGCACGATGCCAATGG	
Stellaris mRNA SLC_7	AGAAGAAGAGCACCTGGTTC	
Stellaris mRNA SLC_8	ACGGACACCGTGAAGATGAT	
Stellaris mRNA SLC_9	TCGTCGATGCTGAACAGGTG	
Stellaris mRNA SLC_10	ACTGTACCTCGTCTGGGAAG	
Stellaris mRNA SLC_11	CTTTTGCTGCGCATGATGTC	
Stellaris mRNA SLC_12	AAGATGGAGGGCTCGATGTC	
Stellaris mRNA SLC_14	TTCTGCATGTCGTACAGGTC	
Stellaris mRNA SLC_15	ACAGAGAACCAGGTGAGGAG	
Stellaris mRNA SLC_16	GAAGTCGGTGTAGAACACGG	
Stellaris mRNA SLC_17	TCGCCTTCGAAGATGACCTG	
Stellaris mRNA SLC_18	TGGCGGCATAAATGACCAGG	
Stellaris mRNA SLC_19	AGGGCTGAACAAATAGCACC	
Stellaris mRNA SLC_20	CGTAGTTGTCCAAGTACTTC	
Stellaris mRNA SLC_21	TAGATCACCCTGACGCTCAG	
Stellaris mRNA SLC_22	CGACGTAGACGTTGGGAAAC	
Stellaris mRNA SLC_23	ATGGTGCTGATGGTGACCAT	
Stellaris mRNA SLC_24	GAGATGCTCATGGAGACGAT	
Stellaris mRNA SLC_25	ACTGCTTGATGTCATGGTAC	
Stellaris mRNA SLC_26	CAAACCCTCGCTTGGAGTTC	
Stellaris mRNA SLC_27	AGGATCTGCGAGATGTACAC	
Stellaris mRNA SLC_28	TTGGGATAGATCACCAGGAA	
Stellaris mRNA SLC_29	AACGGGGAAGACAGGCCTTT	
siRNA circSLC45A4 1	GTTGCAGATTGATCATCCA-dTdT	siRNAs against human circSLC45A4
siRNA circSLC45A4 2	AATACTGTTGCAGATTGATCAT-dTdT	
siRNA circSLC45A4 3	GATTGATCATCCAGTCACTC-dTdT	
shRNA circSlc45a4 1	GATCGTTACAGATAGATCGTCTATTCAAGAGA TAGACGATCTATCTGTAACTTTTTGAAA	shRNAs against mouse circSlc45a4
shRNA circSlc45a4 2	GATCGTATACTGTTACAGATAGATCGTTTCAAGAGAACG ATCTATCTGTAACAGTATACTTTTTGAAA	
shRNA circSlc45a4 3	GATCGATAGATCGTCTAGCCGCTCTTCAAGAG AGAGCGGCTAGACGATCTATCTTTTTGAAA	
shRNA mRNA Slc45a4 1	GATCGGGACGACACCTTGCTTGATAACTCGAGTTATC AAGCAAGGTGTCGTCCTTTTTGAAA	shRNAs against mouse mRNA Slc45a4
shRNA mRNA Slc45a4 2	GATCGGCTGGTACATTCCACACTGGACTCGAGTCCA GTGTGGAATGTACCAGCTTTTTGAAA	
Scrambled siRNA	CAGATTCGCGAATGTACGTTT-dTdT	

Northern probe against ZNF609	CCCTCCGGAGGCTCCACTGCTCAAGGACATCTTAGAGTC AACGTCCCACCTCAAGATTCAAGGCTCTTCCTTGCCCCCA GCTTTCCTATTTTCAGCTTCAGCTATGTTCTCAGACCTGCC ACATTGGTCAGTACATGCCCAGTGGACAACATCATTGCTC TTTTCAGACTTGACTTTCTTTAGCAGGCGACACTCATTCTC CCCTTCCTCTGGCTCCGGTTTAGTTCCCAAAGGATTGAGC GCTGCCCCAGGCTCAATAGCTATACTCCCAAGTGGCTCC ACAGCTCCTGTATCCACCCCTGCACTTCCATCATACTGAC CACCCCGTCCTCCCAAGGGAAC	
Northern probe against XPO1	GGAAAATGTGATAAAAACAAGGTGGAAGATTCTTCCAAGG AACCAGTGCGAAGGCAAAGAATGGCTCAAGAAGTACTGA CACATTTAAAGGAGCATCCTGATGCTTGGACAAGAGT	
human circSLC45A4	FWD: GAGGGGTCCATAGACCGAAT REV: GGTAATCCCCATCGAGTGAC	
human mRNA SLC45A4	FWD: CTTTTCCTTAACGGCTCTGC REV: AGAGAAGGCGTGGATGTTGA	
human pre-mRNA SLC45A4	FWD: GAGGGGTCCATAGACCGAAT REV: TGCCTCCAGAGGATCAGAG	
human NOTCH1	FWD: TGAATGGCGGGAAGTGTGAA REV: CACAGCTGCAGGCATAGTCT	
human PAX6	FWD: ACAGAGGTCAGGCTTCGCTA REV: GGGTTGGTGTGTGAGAGCAA	
human ASCL1	FWD: AGGAGCTTCTCGACTTCACCA REV: ACTAAAGATGCAGGTTGTGCG	
human TUBB3	FWD: GGCCTCTTCTCACAAGTACG REV: CCACTCTGACCAAAGATGAAA	
human GABBR1	FWD: GCCAATACCCGCAGCATTTC REV: TAGTTGAAGTCCTCCAGGCG	
human VGLUT1	FWD: ATCCTGCAGGGGTTGGTAGA REV: GGAACCACAAAAGGCTGTCG	
human ACTB	FWD: CCAACCGCGAGAAGATGA REV: CCAGAGGCGTACAGGGATAG	
human GAPDH	FWD: AAGGTGAAGGTCGGAGTCAAC REV: GGGGTCATTGATGGCAACAATA	
human MT-CO2	FWD: ACGCATCCTTTACATAACAGAC REV: GCCAATTGATTTGATGGTAAGG	
human XIST	FWD: GACACAAGGCCAACGACCTA REV: TCGCTTGGGTCCTCTATCCA	
human NEAT1	FWD: AGTTAAGGCGCCATCCTCAC REV: AGCACTGCCACCTGGAAAAT	
human CCND1	FWD: GATGCCAACCTCCTCAACGA REV: GGAAGCGGTCCAGGTAGTTC	
human PRKCA	FWD: ACCATGGCTGACGTTTTCCC REV: CAAACTTGGCACTGGAAGCC	
human PLXNB2	FWD: GGCGGACGTGCAGAGAATC REV: AGGTGGTTCAGCTCTTTCTCG	qPCR primer
human PLXNA2	FWD: GACAGGCACAGCATCCATGA REV: GTCAAACACGAACTGGGGGT	
human PLXNA3	FWD: AACTGCTCTACGCCAAGGAC REV: CAAGAGGCATCTCGGTCCAG	
human TIMM17B	FWD: TTGTGGATGATTGCGGTGGA REV: CAAGGGAGCTTCAGACAGCA	
human TMEM26	FWD: TGACCTGGCAGTACAGAACG REV: ACGCTGATTCCGATGTTCCA	
human NTRK2	FWD: TTACGGTTTGTCACCCGACC REV: GTGCTTGGTTCAGCTCTTGC	
human ELMOD2	FWD: AGAGCTACTCATGAAGCTTTGGA REV: ATGCCCATGCCTCTGAAGTC	
mouse circSlc45a4	FWD: AGACCCAAGAGGAGACCACA REV: TCGAGCGGCTAGACGATCTATC	
mouse mRNA Slc45a4	FWD: TGGGTGTTGTCTCGATGAGC REV: ACAGAGGCCACTATGGGGAT	
mouse Unc5b	FWD: TGGAGGACACACCTGTAGCA REV: TCATGGAGGGAGAGGCGTAG	
mouse Foxp2	FWD: GTTCATCCGTCAGTGCCCG REV: GTGAGGATCGAGCTTCAACAGA	
mouse Notch2	FWD: CAGCACTGTGACAGCCCTTA REV: TATTCCGCTCACAGGTGCTC	
mouse Parm1	FWD: GCCCAAGGAGAAATCACCCC REV: CAGCAACACTACGGCGATGA	
mouse Itga9	FWD: CTGGAGCACTTCCACGACAA REV: ATCTCCGGTCAGGGTTGGTA	
mouse Cux2	FWD: TCCAGGGTCTTCCGGACATC REV: CCGAGGCGACAGAACTAAGC	
mouse Scrt1	FWD: ATTCTCTTCGGCAGACCTCG REV: TACATAGGCTCCGGGGATGG	
mouse Nkiras2	FWD: CACCTCAGCGGGAGTGC REV: ATGTCCTCCTGGGTCTCGAT	
mouse Trank1	FWD: AGCCTTATGAAGCTGCTCGG	

	REV: TGACAACAGTAGAAATGAATGCCA
	FWD: GGTCCAGGCCATCAAGAGTTT
mouse Nr0b1	REV: CCGGATGTGCTCAGTAAGGAT
	FWD: ACAAGCGAACCATCGAAAAG
mouse Eef1a	REV: GTCTCGAATTTCCACAGGGA

Name	Sequence 5'→ 3'	Purpose
3D-seq_RT_1	[PHO]AGGCCAGAGCATTCGCCTGCGCCTAATNNNNNNNNTTTTTTTTTTTTTTTVN	3D-seq RT primer with barcode, UMI and ligation linker
3D-seq_RT_2	[PHO]AGGCCAGAGCATTCGAGTATCCAGGACNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_3	[PHO]AGGCCAGAGCATTCGGTCCTACTTGAGNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_4	[PHO]AGGCCAGAGCATTCGATCTCTCGGTAGNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_5	[PHO]AGGCCAGAGCATTCGGGCGGATTCTTCNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_6	[PHO]AGGCCAGAGCATTCGGATGAGTAACTGNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_7	[PHO]AGGCCAGAGCATTCGAATGTTACGTTGNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_8	[PHO]AGGCCAGAGCATTCGACCTGCGTATTGNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_9	[PHO]AGGCCAGAGCATTCGCGTGTGAATGGCNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_10	[PHO]AGGCCAGAGCATTCGAACCACCTCCTTNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_11	[PHO]AGGCCAGAGCATTCGCAGTCCACAGCCNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_12	[PHO]AGGCCAGAGCATTCGTGTCGAGCTGGCNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_13	[PHO]AGGCCAGAGCATTCGTGACTGGTATGANNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_14	[PHO]AGGCCAGAGCATTCGCGGCAAGATTCANNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_15	[PHO]AGGCCAGAGCATTCGATAGAACGATGTNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_16	[PHO]AGGCCAGAGCATTCGCTGAGTCGCGGCNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_17	[PHO]AGGCCAGAGCATTCGTGAGAGTTCGCTNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_18	[PHO]AGGCCAGAGCATTCGGTGCGGAGGCATNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_19	[PHO]AGGCCAGAGCATTCGTCTACGTGAACANNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_20	[PHO]AGGCCAGAGCATTCGCTTAGGAACGCANNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_21	[PHO]AGGCCAGAGCATTCGACGACGTCTCATNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_22	[PHO]AGGCCAGAGCATTCGGTTATTAGCGTTNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_23	[PHO]AGGCCAGAGCATTCGCCTTGTACTTAGNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_24	[PHO]AGGCCAGAGCATTCGCCGTAAGATGATNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_25	[PHO]AGGCCAGAGCATTCGTCTGGCCGCCACNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_26	[PHO]AGGCCAGAGCATTCGATACCGGAGACTNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_27	[PHO]AGGCCAGAGCATTCGATTATTCTAGCCNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_28	[PHO]AGGCCAGAGCATTCGGAACACAATCATNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_29	[PHO]AGGCCAGAGCATTCGGCGGTCTCCACTNNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_RT_30	[PHO]AGGCCAGAGCATTCGAAGGTCGTGAGANNNNNNNNTTTTTTTTTTTTTTTVN	
3D-seq_bc2_1	CAGACGTGTGCTCTTCCGATCTATGCTACCAGACCCCCATGATCGTCCGA	3D-seq ligation oligonucleotide, containing barcode 2, the ligation linker and the PCR handle
3D-seq_bc2_2	CAGACGTGTGCTCTTCCGATCTAATCTAAGGATTCCCCATGATCGTCCGA	
3D-seq_bc2_3	CAGACGTGTGCTCTTCCGATCTAACAGTAAGCCTCCCCATGATCGTCCGA	
3D-seq_bc2_4	CAGACGTGTGCTCTTCCGATCTAGTTGCAGATGAGCCCATGATCGTCCGA	
3D-seq_bc2_5	CAGACGTGTGCTCTTCCGATCTACGAACGTCAGACCCCATGATCGTCCGA	
3D-seq_bc2_6	CAGACGTGTGCTCTTCCGATCTATCGTCCTATGGCCCCATGATCGTCCGA	
3D-seq_bc2_7	CAGACGTGTGCTCTTCCGATCTACTCAACATAATGCCCATGATCGTCCGA	
3D-seq_bc2_8	CAGACGTGTGCTCTTCCGATCTATGCGAATACCGACCCATGATCGTCCGA	
3D-seq_bc2_9	CAGACGTGTGCTCTTCCGATCTAAAGTCGGAGTCCCCCATGATCGTCCGA	
3D-seq_bc2_10	CAGACGTGTGCTCTTCCGATCTAGACATTCCGATTCCCATGATCGTCCGA	
3D-seq_bc2_11	CAGACGTGTGCTCTTCCGATCTACTTACTGAACGGCCCATGATCGTCCGA	
3D-seq_bc2_12	CAGACGTGTGCTCTTCCGATCTACCAGGCTGTTGGCCCATGATCGTCCGA	
3D-seq_bc2_13	CAGACGTGTGCTCTTCCGATCTACACTCCACCGCTCCCATGATCGTCCGA	
3D-seq_bc2_14	CAGACGTGTGCTCTTCCGATCTAACCGACGCCGTACCCATGATCGTCCGA	
3D-seq_bc2_15	CAGACGTGTGCTCTTCCGATCTATATCGGTATTAGCCCATGATCGTCCGA	
3D-seq_bc2_16	CAGACGTGTGCTCTTCCGATCTAGTCCTACTGGATCCCATGATCGTCCGA	
3D-seq_bc2_17	CAGACGTGTGCTCTTCCGATCTATCTATGTCACTTCCCATGATCGTCCGA	
3D-seq_bc2_18	CAGACGTGTGCTCTTCCGATCTATCCAGACTTCGTCCCATGATCGTCCGA	
3D-seq_bc2_19	CAGACGTGTGCTCTTCCGATCTATAGGTCACTCGGCCCATGATCGTCCGA	
3D-seq_bc2_20	CAGACGTGTGCTCTTCCGATCTATTGTAGGCAGCTCCCATGATCGTCCGA	
3D-seq_bc2_21	CAGACGTGTGCTCTTCCGATCTATGGATAGCCATCCCCATGATCGTCCGA	
3D-seq_bc2_22	CAGACGTGTGCTCTTCCGATCTAAGACCTTGCAGCCCCATGATCGTCCGA	
3D-seq_bc2_23	CAGACGTGTGCTCTTCCGATCTACAACAACGCTTACCCATGATCGTCCGA	
3D-seq_bc2_24	CAGACGTGTGCTCTTCCGATCTAGCTAGGATCTTACCCATGATCGTCCGA	
3D-seq_bc2_25	CAGACGTGTGCTCTTCCGATCTAAGTGGTGAGGCACCCATGATCGTCCGA	
3D-seq_bc2_26	CAGACGTGTGCTCTTCCGATCTAGAGCCGGTTATTCCCATGATCGTCCGA	
3D-seq_bc2_27	CAGACGTGTGCTCTTCCGATCTACACCTGCTATCTCCCATGATCGTCCGA	
3D-seq_bc2_28	CAGACGTGTGCTCTTCCGATCTAAGATGGCTAGTGCCCATGATCGTCCGA	
3D-seq_bc2_29	CAGACGTGTGCTCTTCCGATCTATCAATGCAGACGCCCATGATCGTCCGA	
3D-seq_bc2_30	CAGACGTGTGCTCTTCCGATCTACCGCAGACATGCCCCATGATCGTCCGA	
3D-seq_bc2_31	CAGACGTGTGCTCTTCCGATCTAGACGGACCCATGATCGTCCGA	
3D-seq_bc2_32	CAGACGTGTGCTCTTCCGATCTAGCGTCTACCCATGATCGTCCGA	
3D-seq_bc2_33	CAGACGTGTGCTCTTCCGATCTAGCCAGCCACCCATGATCGTCCGA	
3D-seq_bc2_34	CAGACGTGTGCTCTTCCGATCTACTGTAGAATCCCATGATCGTCCGA	
3D-seq_bc2_35	CAGACGTGTGCTCTTCCGATCTAACTACCATGCCCCATGATCGTCCGA	

3D-seq_bc2_36	CAGACGTGTGCTCTTCCGATCTAATCCAGGACAGTCCCATGATCGTCCGA
3D-seq_bc2_37	CAGACGTGTGCTCTTCCGATCTACCAAGATCGCAATCCCATGATCGTCCGA
3D-seq_bc2_38	CAGACGTGTGCTCTTCCGATCTATCATGAAGGTCACCCATGATCGTCCGA
3D-seq_bc2_39	CAGACGTGTGCTCTTCCGATCTACGATCAGACGGCTCACCCATGATCGTCCGA
3D-seq_bc2_40	CAGACGTGTGCTCTTCCGATCTACATAGGCTGAGCCCCATGATCGTCCGA
3D-seq_bc2_41	CAGACGTGTGCTCTTCCGATCTACTGGTTAACTAGCCCATGATCGTCCGA
3D-seq_bc2_42	CAGACGTGTGCTCTTCCGATCTATTGGAATCTTACCCCATGATCGTCCGA
3D-seq_bc2_43	CAGACGTGTGCTCTTCCGATCTAGGCACTAGTTCTCCCATGATCGTCCGA
3D-seq_bc2_44	CAGACGTGTGCTCTTCCGATCTAGAGTGACTAGACCCATGATCGTCCGA
3D-seq_bc2_45	CAGACGTGTGCTCTTCCGATCTAACGAAGGAACTCCCATGATCGTCCGA
3D-seq_bc2_46	CAGACGTGTGCTCTTCCGATCTATATGTTCGTGCGACCCATGATCGTCCGA
3D-seq_bc2_47	CAGACGTGTGCTCTTCCGATCTACCAATTACCTGCACCCATGATCGTCCGA
3D-seq_bc2_48	CAGACGTGTGCTCTTCCGATCTACAACATCAAGGCGCCCATGATCGTCCGA
3D-seq_bc2_49	CAGACGTGTGCTCTTCCGATCTAGATACATCCGGCATCCCATGATCGTCCGA
3D-seq_bc2_50	CAGACGTGTGCTCTTCCGATCTATCCGCCACGACTATGCCCATGATCGTCCGA
3D-seq_bc2_51	CAGACGTGTGCTCTTCCGATCTAACCAGTTGTCCTCGACCCATGATCGTCCGA
3D-seq_bc2_52	CAGACGTGTGCTCTTCCGATCTAGTCTTGTCGTGCCGACCCATGATCGTCCGA
3D-seq_bc2_53	CAGACGTGTGCTCTTCCGATCTATAACAGCGAATGGACACCCATGATCGTCCGA
3D-seq_bc2_54	CAGACGTGTGCTCTTCCGATCTACTAGGCTGCCTCAATGCCCATGATCGTCCGA
3D-seq_bc2_55	CAGACGTGTGCTCTTCCGATCTAACTCGACATCCCCATGATCGTCCGA
3D-seq_bc2_56	CAGACGTGTGCTCTTCCGATCTAGCCGAGTAAGCCCATGATCGTCCGA
3D-seq_bc2_57	CAGACGTGTGCTCTTCCGATCTACCAGTTGGTACCCATGATCGTCCGA
3D-seq_bc2_58	CAGACGTGTGCTCTTCCGATCTATCTGGACCGCCCATGATCGTCCGA
3D-seq_bc2_59	CAGACGTGTGCTCTTCCGATCTACAAGCGACCCCCATGATCGTCCGA
3D-seq_bc2_60	CAGACGTGTGCTCTTCCGATCTACTCCGAGGACCCATGATCGTCCGA
3D-seq_bc2_61	CAGACGTGTGCTCTTCCGATCTATTGCAGTCCCCATGATCGTCCGA
3D-seq_bc2_62	CAGACGTGTGCTCTTCCGATCTACACTTGAGCCCATGATCGTCCGA
3D-seq_bc2_63	CAGACGTGTGCTCTTCCGATCTACGGAATCTCCCATGATCGTCCGA
3D-seq_bc2_64	CAGACGTGTGCTCTTCCGATCTAACAGAAGCCCATGATCGTCCGA
3D-seq_bc2_65	CAGACGTGTGCTCTTCCGATCTATTAACGCCCCATGATCGTCCGA
3D-seq_bc2_66	CAGACGTGTGCTCTTCCGATCTAAGCTTGCCCCATGATCGTCCGA
3D-seq_bc2_67	CAGACGTGTGCTCTTCCGATCTAGGTCAGTCCCATGATCGTCCGA
3D-seq_bc2_68	CAGACGTGTGCTCTTCCGATCTATAGTCCCCCATGATCGTCCGA
3D-seq_bc2_69	CAGACGTGTGCTCTTCCGATCTAAGAACACCCATGATCGTCCGA
3D-seq_bc2_70	CAGACGTGTGCTCTTCCGATCTAGATTGTCATCTACCCATGATCGTCCGA
3D-seq_bc2_71	CAGACGTGTGCTCTTCCGATCTAATGCCTCCAATACCCATGATCGTCCGA
3D-seq_bc2_72	CAGACGTGTGCTCTTCCGATCTAACCGTACATTCACCCATGATCGTCCGA
3D-seq_bc2_73	CAGACGTGTGCTCTTCCGATCTACCGGTGACGACTGACCCATGATCGTCCGA
3D-seq_bc2_74	CAGACGTGTGCTCTTCCGATCTATGAGTTCTTACGTGCCCATGATCGTCCGA
3D-seq_bc2_75	CAGACGTGTGCTCTTCCGATCTACTATCTACGCCACGGACCCATGATCGTCCGA
3D-seq_bc2_76	CAGACGTGTGCTCTTCCGATCTACTTCCGACTCCACCCATGATCGTCCGA
3D-seq_bc2_77	CAGACGTGTGCTCTTCCGATCTAGTGAGCGTGCAGCCCATGATCGTCCGA
3D-seq_bc2_78	CAGACGTGTGCTCTTCCGATCTATGTTCGAACTGTCCCATGATCGTCCGA
3D-seq_bc2_79	CAGACGTGTGCTCTTCCGATCTAGGATTCTTCTAACCCATGATCGTCCGA
3D-seq_bc2_80	CAGACGTGTGCTCTTCCGATCTAACTGCACTCAATCCCATGATCGTCCGA
3D-seq_bc2_81	CAGACGTGTGCTCTTCCGATCTACACGATACGCAACCCATGATCGTCCGA
3D-seq_bc2_82	CAGACGTGTGCTCTTCCGATCTAAATTCCGGAAGACCCATGATCGTCCGA
3D-seq_bc2_83	CAGACGTGTGCTCTTCCGATCTATCACCGGTCCAACCCATGATCGTCCGA
3D-seq_bc2_84	CAGACGTGTGCTCTTCCGATCTAAAGACATTGGTGCCCATGATCGTCCGA
3D-seq_bc2_85	CAGACGTGTGCTCTTCCGATCTATCTGCATTAGCCCCATGATCGTCCGA
3D-seq_bc2_86	CAGACGTGTGCTCTTCCGATCTAGACCAGTATTCCCCATGATCGTCCGA
3D-seq_bc2_87	CAGACGTGTGCTCTTCCGATCTACGTCGCAATGACCCATGATCGTCCGA
3D-seq_bc2_88	CAGACGTGTGCTCTTCCGATCTAGTGGACGTACCCCCATGATCGTCCGA
3D-seq_bc2_89	CAGACGTGTGCTCTTCCGATCTAGCCATTGCTGGTCGCCCATGATCGTCCGA
3D-seq_bc2_90	CAGACGTGTGCTCTTCCGATCTACAAGGCCTACCATCCCCATGATCGTCCGA
3D-seq_bc2_91	CAGACGTGTGCTCTTCCGATCTATGTAGCATGCGTTACCCATGATCGTCCGA
3D-seq_bc2_92	CAGACGTGTGCTCTTCCGATCTAATCGCAAGCGGCTGCCCATGATCGTCCGA
3D-seq_bc2_93	CAGACGTGTGCTCTTCCGATCTATTGCTCCACCACGACCCATGATCGTCCGA
3D-seq_bc2_94	CAGACGTGTGCTCTTCCGATCTATTATGTTCGGCCATAACCCATGATCGTCCGA
3D-seq_bc2_95	CAGACGTGTGCTCTTCCGATCTAAATCCTACAGGCATGACCCATGATCGTCCGA
3D-seq_bc2_96	CAGACGTGTGCTCTTCCGATCTACGCCTCGACTCCCATGATCGTCCGA

P7_N701	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	3D-seq PCR primer, based on Illumina's P7 primer, including P7 site and i7 multiplex
P7_N702	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
P7_N703	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
P7_N704	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
P7_N705	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
P7_N706	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
P7_N707	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
P7_N710	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
P7_N711	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	

P7_N712	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTGACTGGAGTTCAGACGTGTGCT CTTCCGATCT	barcod es
P7_N714	CAAGCAGAAGACGGCATACGAGATTCATGAGCGTGACTGGAGTTCAGACGTGTGCT CTTCCGATCT	
P7_N715	CAAGCAGAAGACGGCATACGAGATCCTGAGATGTGACTGGAGTTCAGACGTGTGCT CTTCCGATCT	

10.2 Definition of used sequences and identifiers

Name	Chromosome coordinates (if applicable)	Database and identifier
circSLC45A4 (human)	chr8:142,264,087-142,264,728 (hg19)	circbase hsa_circ_0001829
circSlc45a4 (mouse)	chr15: 73,435,871 – 73,436,506 (mm9)	circbase mmu_circ_0005786
mRNA SLC45A4 (human)	chr8: 142,217,265 – 142,264,728 (hg19)	RefSeq NM_001286646
mRNA Slc45a4 (mouse)	chr15: 73,407,854 – 73,455,174 (mm9)	RefSeq NM_001033219

10.3 Used marker genes for Bseq-SC analysis

p_val	avg_logFC	pct.1	pct.2	p_val_adj	cluster	gene	marker for	PMID/title/DOI
1.8E-199	1.48	0.598	0.125	3E-195	0	Nrp1	migrating excitatory neurons	28507510
1.5E-121	1.02	0.468	0.114	2.5E-117	0	Rnd2		Regulation of cerebral cortex development by Rho GTPases: insights from in vivo studies
2.13E-78	0.77	0.265	0.051	3.5E-74	0	Unc5d		The cortical subventricular zone-specific molecule Svet1 is part of the nuclear RNA coded by the putative netrin receptor gene Unc5d and is expressed in multipolar migrating cells
1.2E-223	1.30	0.815	0.237	1.9E-219	1	Fezf2	subcortical projection neurons	The Fezf2–Ctip2 genetic pathway regulates the fate choice of subcortical projection neurons in the developing cerebral cortex
1.2E-187	1.05	0.852	0.296	2E-183	1	Dab1		Abcam neuron markers guide
1.2E-126	0.84	0.526	0.13	1.9E-122	1	Ldb2		Generating Neuronal Diversity in the Mammalian Cerebral Cortex
1.7E-115	0.85	0.463	0.11	2.8E-111	1	Ndrg1		
2.7E-105	0.78	0.472	0.123	4.4E-101	1	Rspo3		mousebrain.org
1.83E-86	0.66	0.476	0.145	3.01E-82	1	Hspa12a		
2.92E-82	0.72	0.439	0.131	4.81E-78	1	Sla		mousebrain.org
9.05E-82	0.70	0.57	0.212	1.49E-77	1	Nin		
1.99E-68	0.70	0.303	0.076	3.28E-64	1	Oma1		
1E-251	1.09	0.45	0.031	1.7E-247	2	Hs3st4	Excitatory neurons	mousebrain.org
8.7E-243	1.28	0.65	0.1	1.4E-238	2	Sybu		
8.6E-230	1.34	0.681	0.119	1.4E-225	2	Tle4		Corticothalamic projection neuron development beyond subtype specification: Fog2 and intersectional controls regulate intra-class neuronal diversity
2E-163	0.99	0.573	0.113	3.2E-159	2	Sh3gl2		
1.3E-130	1.04	0.467	0.09	2.1E-126	2	Kitl		Tbr1 regulates regional and laminar identity of postmitotic neurons in developing neocortex
6.2E-127	0.97	0.427	0.077	1E-122	2	Ppp1r1b		mousebrain.org

1.23E-91	0.94	0.412	0.097	2.02E-87	2	Sparcl1		
1.26E-75	0.96	0.379	0.1	2.08E-71	2	Lpl		mousebrain.org
0	1.66	0.477	0.01	0	3	Nxph1	Cortical interneurons	Single-cell RNA sequencing identifies distinct mouse medial ganglionic eminence cell types
0	1.53	0.451	0.01	0	3	Lhx6		mousebrain.org
8.7E-268	1.37	0.364	0.01	1.4E-263	3	Erbb4		Single-cell RNA sequencing identifies distinct mouse medial ganglionic eminence cell types
4.2E-244	2.39	0.345	0.011	6.9E-240	3	Sst		Single-cell RNA sequencing identifies distinct mouse medial ganglionic eminence cell types
2.8E-236	1.72	0.527	0.057	4.7E-232	3	Mafb		Single-cell RNA sequencing identifies distinct mouse medial ganglionic eminence cell types
7.1E-170	1.07	0.278	0.013	1.2E-165	3	Rpp25		
2.7E-148	2.08	0.323	0.03	4.5E-144	3	Npy		Single-cell RNA sequencing identifies distinct mouse medial ganglionic eminence cell types
8E-147	1.26	0.41	0.057	1.3E-142	3	Rph3a		
9.9E-145	1.08	0.254	0.015	1.6E-140	3	Neto1		
2.8E-136	1.27	0.384	0.053	4.5E-132	3	Cbfa2t3		
1.6E-127	1.10	0.312	0.034	2.7E-123	3	Gria4		
6.4E-108	1.11	0.321	0.046	1.1E-103	3	Slc6a1		
6.9E-108	1.18	0.38	0.068	1.1E-103	3	Cux2		
7.38E-96	1.05	0.371	0.074	1.21E-91	3	Dlgap1		
0	2.39	0.846	0.061	0	4	Eomes	Basal progenitor / cortical intermediate progenitor	Pax6, Tbr2, and Tbr1 are expressed sequentially by radial glia, intermediate progenitor cells, and postmitotic neurons in developing neocortex
5.2E-273	1.56	0.601	0.062	8.5E-269	4	Mfap4		Developmental Emergence of Adult Neural Stem Cells as Revealed by Single-Cell Transcriptional Profiling
6.5E-228	1.84	0.603	0.085	1.1E-223	4	Neurog2		Neurog2 Simultaneously Activates and Represses Alternative Gene Expression Programs in the Developing Neocortex
1.8E-179	1.28	0.555	0.087	2.9E-175	4	Sstr2		Developmental Emergence of Adult Neural Stem Cells as Revealed by Single-Cell Transcriptional Profiling
5E-125	0.94	0.265	0.022	8.3E-121	4	Btbd17		
0	2.26	0.732	0.073	0	5	Top2a	Neural Stem/Progenitor cells	Distinct expression patterns for type II topoisomerases IIA and IIB in the early foetal human telencephalon.
0	2.11	0.79	0.063	0	5	2810417H13Rik		
0	1.82	0.653	0.053	0	5	Mki67		
3.4E-307	1.83	0.697	0.083	5.5E-303	5	Rrm2		
2.4E-281	1.71	0.558	0.047	3.9E-277	5	Nusap1		Developmental Emergence of Adult Neural Stem Cells as Revealed by Single-Cell Transcriptional Profiling

6.7E-251	1.55	0.498	0.041	1.1E-246	5	Ckap2l		Radmis, a Novel Mitotic Spindle Protein that Functions in Cell Division of Neural Progenitors
1.6E-248	1.76	0.538	0.052	2.6E-244	5	Prc1		
6.6E-224	1.63	0.469	0.041	1.1E-219	5	Cenpe		
5.4E-249	1.46	0.699	0.098	8.9E-245	6	Satb2	Upper layer pyramidal neurons	Satb2 Is a Postmitotic Determinant for Upper-Layer Neuron Specification in the Neocortex
4.4E-244	1.43	0.758	0.119	7.3E-240	6	913002 4F11Rik		
2E-112	0.89	0.403	0.064	3.4E-108	6	Plxna4		Pyramidal Neurons Grow Up and Change Their Mind
3.55E-89	1.10	0.374	0.073	5.83E-85	6	Ppp1r14c		
0	2.13	0.544	0.021	0	7	Isl1	Maturing + adult cortical interneurons	
2.1E-280	2.49	0.507	0.025	3.4E-276	7	Ebf1		
2.7E-231	1.59	0.405	0.017	4.5E-227	7	Zfp503		
3.67E-90	1.00	0.294	0.033	6.05E-86	7	Dlx6		Dlx5 and Dlx6 Regulate the Development of Parvalbumin-Expressing Cortical Interneurons
8.89E-89	1.08	0.277	0.03	1.46E-84	7	Vstm2a		
1.15E-43	1.06	0.267	0.059	1.9E-39	7	Foxp2		FoxP2 regulates neurogenesis during embryonic cortical development.
0	2.04	0.735	0.035	0	8	Aldoc	Radial precursor cells	Developmental Emergence of Adult Neural Stem Cells as Revealed by Single-Cell Transcriptional Profiling
0	1.67	0.636	0.033	0	8	Gas1		
0	1.55	0.591	0.024	0	8	Mfge8		
9.3E-305	1.43	0.505	0.019	1.5E-300	8	Rgcc		
4E-286	1.88	0.773	0.083	6.6E-282	8	Ttyh1		
2.5E-257	1.78	0.57	0.04	4.1E-253	8	Mt2		Developmental Emergence of Adult Neural Stem Cells as Revealed by Single-Cell Transcriptional Profiling
4.5E-255	1.47	0.622	0.05	7.4E-251	8	Ppap2b		
3.1E-232	1.45	0.543	0.04	5E-228	8	Slc1a3		Developmental Emergence of Adult Neural Stem Cells as Revealed by Single-Cell Transcriptional Profiling
6.3E-227	1.70	0.704	0.086	1E-222	8	Ddah1		
3E-222	1.40	0.646	0.067	4.9E-218	8	Sox9		Developmental Emergence of Adult Neural Stem Cells as Revealed by Single-Cell Transcriptional Profiling
1.3E-172	1.61	0.44	0.037	2.2E-168	8	Mt3		Developmental Emergence of Adult Neural Stem Cells as Revealed by Single-Cell Transcriptional Profiling
2.33E-81	1.43	0.297	0.033	3.83E-77	9	63	Neural progenitors	63 Controls the Neural Progenitor Status in the Murine CNS
0	2.76	0.904	0.029	0	10	Reln	Cajal-Retzius cells	
0	2.30	0.606	0.01	0	10	Snhg11		
0	2.16	0.819	0.024	0	10	Cacna2d2		
0	1.93	0.723	0.003	0	10	Lhx5		LIM-homeobox gene Lhx5 is required for normal

								development of Cajal-Retzius cells.
0	1.92	0.638	0.002	0	10	Lhx1		LIM-homeobox gene Lhx5 is required for normal development of Cajal-Retzius cells.
0	1.91	0.67	0.002	0	10	1500016L03Rik		
0	1.68	0.66	0.012	0	10	Nr2f2		
0	1.47	0.574	0.002	0	10	RP24-351J24.2		
6.4E-270	1.39	0.596	0.014	1.1E-265	10	Rcan2		Gene expression analysis of the embryonic subplate
1.4E-248	1.97	0.617	0.018	2.2E-244	10	Calb2		
6.4E-153	1.40	0.426	0.015	1.1E-148	10	Cdkn1a		
0	3.17	0.967	0.006	0	11	Alas2	erythrocytes	Hemoglobin mRNA Changes in the Frontal Cortex of Patients with Neurodegenerative Diseases
0	2.71	0.767	0.012	0	11	Car2		
0	2.31	0.9	0.001	0	11	Hbb-bh1		
0	2.08	0.683	0	0	11	Gypa		
0	1.89	0.767	0.01	0	11	Slc25a37		
0	1.60	0.8	0.01	0	11	Hbb-y		
0	1.52	0.7	0.003	0	11	Fam46c		
0	1.39	0.5	0	0	11	Trim10		
1.1E-245	1.33	0.5	0.007	1.7E-241	11	Ube2l6		
1.3E-167	1.50	0.633	0.022	2.2E-163	11	Blvrb		
1.6E-158	2.11	0.85	0.05	2.6E-154	11	Fech		
1E-131	1.63	0.55	0.022	1.7E-127	11	Pnpo		
4.44E-23	1.48	0.963	0.658	7.31E-19	12	Gria2	Excitatory neurons	mousebrain.org
2.91E-18	1.68	0.722	0.357	4.79E-14	12	Ptprs		mousebrain.org
5.77E-16	1.68	0.556	0.199	9.5E-12	12	Arpp21		
1.31E-11	1.50	0.63	0.349	2.16E-07	12	Cadm1		mousebrain.org
1.99E-10	1.44	0.352	0.111	3.27E-06	12	Ankrd13b		
8.81E-08	1.28	0.519	0.296	0.00145	12	Ank2		mousebrain.org
0	3.63	0.951	0.002	0	13	P2ry12	microglia	Microglia versus myeloid cell nomenclature during brain inflammation
0	3.45	0.951	0.001	0	13	C1qb		Microglia versus myeloid cell nomenclature during brain inflammation
0	3.37	0.976	0.002	0	13	Ctss		Microglia versus myeloid cell nomenclature during brain inflammation
0	3.36	0.951	0.002	0	13	C1qc		Microglia versus myeloid cell nomenclature during brain inflammation
0	3.33	0.951	0.003	0	13	Cx3cr1		Microglia versus myeloid cell nomenclature during brain inflammation
0	3.31	0.951	0.01	0	13	Rnase4		Microglia versus myeloid cell nomenclature during brain inflammation
0	3.22	0.927	0.001	0	13	C1qa		Microglia versus myeloid cell nomenclature during brain inflammation

0	3.21	0.951	0.002	0	13	Laptm5		Microglia versus myeloid cell nomenclature during brain inflammation
0	3.19	0.951	0.001	0	13	Tyrobp		Microglia versus myeloid cell nomenclature during brain inflammation
0	3.19	0.951	0	0	13	Fcrls		Microglia versus myeloid cell nomenclature during brain inflammation
0	3.17	0.976	0.001	0	13	Csf1r		Microglia versus myeloid cell nomenclature during brain inflammation
0	2.97	0.927	0.002	0	13	Fcer1g		Microglia versus myeloid cell nomenclature during brain inflammation
0	2.94	0.927	0.007	0	13	Sepp1		Microglia versus myeloid cell nomenclature during brain inflammation
2.3E-147	3.44	0.927	0.047	3.8E-143	13	Hexb		Microglia versus myeloid cell nomenclature during brain inflammation
1.2E-131	4.10	0.951	0.057	2E-127	13	Apoe		Microglia versus myeloid cell nomenclature during brain inflammation

10.4 Detailed protocol for a 3D-seq run

All cell pelleting steps are done in Maxymum Recovery tubes.

Store all cell counting images on USB – after RT and after Ligation.

Fixation of tissue with – MeOH

1. Take tissue slice containing boxes from -80°C and let come to room temperature
    a. Will take approx. 30 min
2. Will waiting:
    a. wash grids 2x with 100 % EtOH to remove hydrophobic rests, wash with water and let air dry
    b. Thaw RT primers, Superscript IV buffer, DTT, dNTPs on ice
3. Fix slices in cold MeOH (stored at -20°C), 20 min @ -20°C
4. Dry slices at room temperature and wash 3x with cold PBS/Ribolock (1:80)
5. Fix grid on tissue slice and take pictures with Keyence microscope
    a. Place tissue holding slide with tissue facing up in Keyence microscope, select 4x objective, Color camera and use Multi-Dimensional Capture/ XY-Stitching -> mark outer rims of tissue and let program do the capturing (should take ~2-3 min) **Figure 1**
    b. Take out tissue slide and fix grid on tissue:
        i. place the bottom of the mounting device
        ii. put the grid in the top part of the mounting device and place it on the tissue
        iii. use screwdriver and 4 screws to tightly fix grid on tissue (start with diagonal screws)
        iv. place slice with grid back in Keyence microscope, adjust Brightness and Contrast until you can see the tissue through the grid (usually one click in Image -> Lookup Table -> Auto Shadow & Highlight) **Figure 2**
        v. use XY-stitching mode for taking images

Figure 40: Keyence setting for tissue slice capture

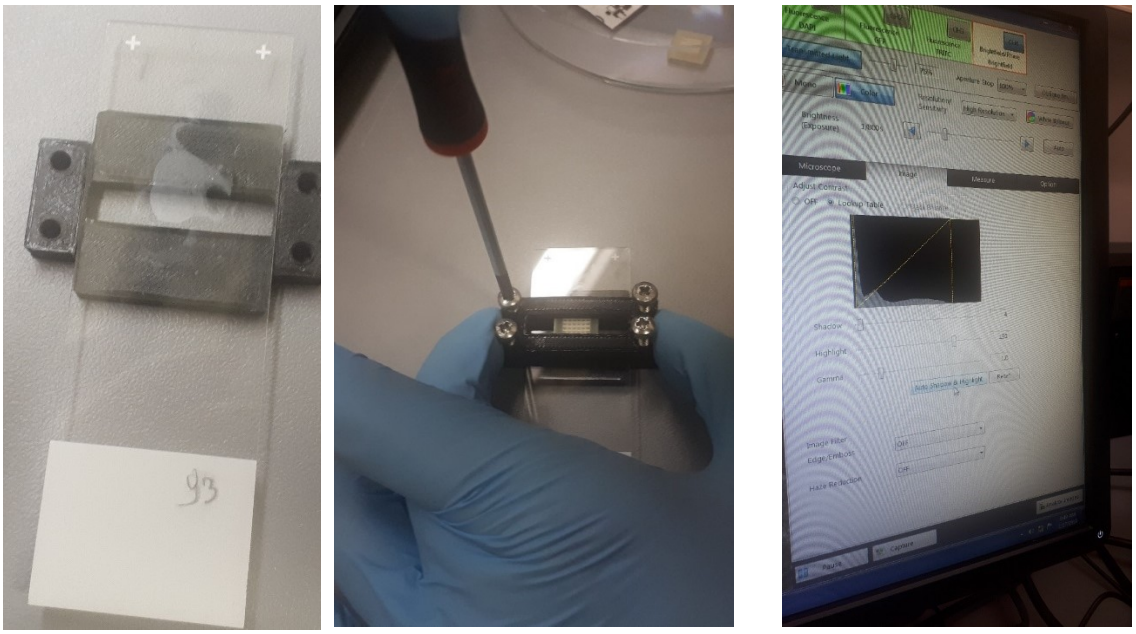


Figure 41: Taking images of tissue with grid on top. Left to right: Bottom part of mounting device with microscope slide that holds tissue; Grid within upper part of mounting device being screwed to the bottom part; Auto Shadow & Brightness adjustment

RT without template switch

Total of 1.5 µl reaction per grid well; total of 30 reactions:

1. Put RT primer in PCR plate (same orientation as you want to have it in the grid)

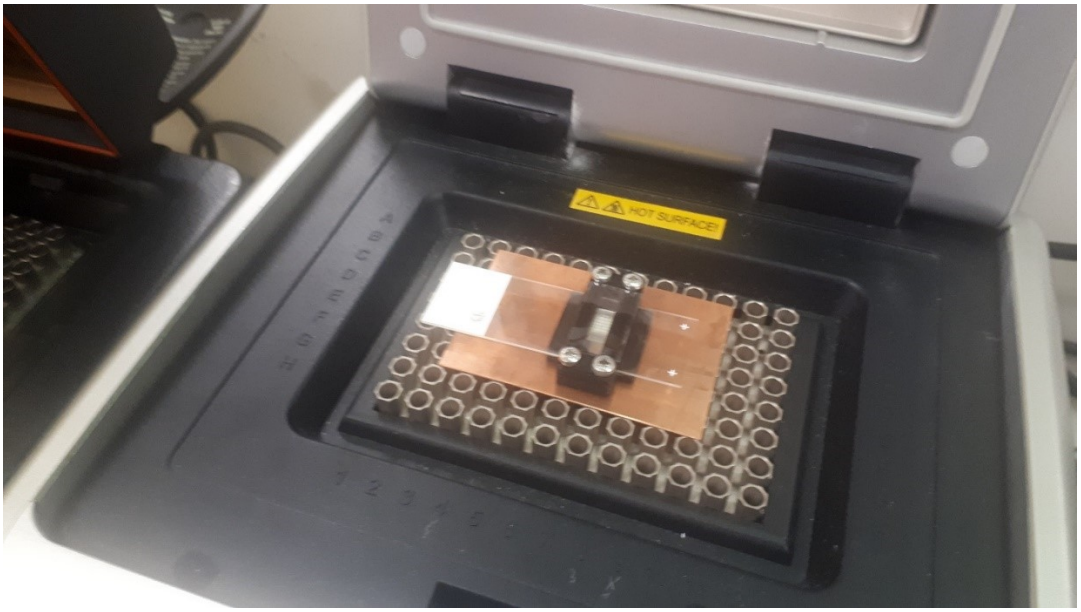
1	7	13	19	25
2	8	14	20	26
3	9	15	21	27
4	10	16	22	28
5	11	17	23	29
6	12	18	24	30

2. Mix remaining Mastermix separately, add 1.25 µl to RT primer in PCR plate

	Per well	For 30 wells = 32x
RT primer 100 µM	0.50	-----
10 mM dNTPs	0.25	8.0
5x Superscript IV Buffer	0.40	12.8
Ribolock	0.25	8.0
100 mM DTT	0.10	3.2
Superscript IV	0.25	8.0

3. Add 1.5 µl of Mastermix to each grid compartment, use capillary tips and LED magnifying glasses
4. Close grid with cut qPCR plate foil seal
5. Incubate: 25°C for 30 min, 37°C for 30 min; 45°C for 20 min (place on copper plate in 96 well cycler), program is stored on silver BIOER cycler -> Christin ->

3D_RT_SPLIT; tighten lid of PCR cycler; use “Block control, highest possible sample volume”, hot lid control: 50°C



6. Meanwhile:
    - prepare ligation plate (see “Ligation of barcode_2”)
    - prepare Papain for tissue digestion 15 min before RT is over
        - aliquot powder of one Worthington Papain Vial (Worthington LK003178) in 3 Eppendorf tubes // or use one aliquot
        - Add 500 µl EBSS
        - Incubate 10 min at 37°C -> cell culture incubator for O_2:CO_2 equilibration, check if color changed from very pink to orange
7. Carefully open PCR cycler (screws could stick to top part, as they are magnetic)
8. Unscrew grid and wash slice carefully with 1xPBS/Ribolock (be fast)
9. Dry surrounding of tissue with Q-tip, circle with hydrophobic pen
10. Prepare single cell suspension:
    - Place microscope slide in plastic box that can be closed with a lid and place a water soaked Kimwipe next to is (to avoid drying out)

- Add 150-300 µl prepared, equilibrated Papain solution to slice and incubate 15 min @ 37˚C, in cell culture incubator, plastic box closed

- wash slice off into Eppendorf tube, if slice parts stick, scratch off with pipette tip
- Completely harvest slice into Eppendorf tube, use up to 300 µl PBS/Ribolock (1:80) + 0.01 %BSA to aid washing off
- Incubate another 15 min @ 37˚C on a shaker, using program F9, 12 rpm
- Spin down cells 3000g, 15 min, 4˚C, resuspend in **320 µl**
- Take out 6 µl and mix with 6 µl Trypan blue for cell counting
- Cell count: _____ cells/ml (DEAD)

Ligation of barcode_2 –10.0 µM

1. Prepare ligation **barcode plate**:
   a. Anneal oCS632 and Split_bc2_**1-96**
   b. Dilute all barcode stocks to 50 µM

	For 5.0 µl 20 µM barcode mix	x97
oCS632 50 µM	2.0	194.0
Split_bc2 50 µM	2.0	---
CutSmart Buffer 10x	0.5	48.5
water	0.5	48.5

   c. Heat to 95°C, 5 min, ramp down to 20°C, -0.1°C/s (Eppendorf Mastercycler X50s: 3D-seq -> Ligation annealing)

2. Ligation reaction, add 5 µl to the annealed 5 µl in the barcode plate

	10 µl Rxn	x97
Cell Solution	3.0	291.0
T4 DNA Ligation Buffer	1.0	97.0
T4 DNA Ligase (10 U/µl final)	0.25	24.25
water	0.75	72.75

a. Incubate 1 h at room temperature, shaking on Eppendorf MixMate 500 rpm

b. Collect all cells (~1000 µl) and add BSA to 0.01 %

c. Centrifuge at 3000g, 4°C, 15 min, take off supernatant (but leave 20-40 µl to not disturb the tiny cell pellet), resuspend in 960 µl of PBS/BSA + Ribolock

d. Count cells: _____ DEAD cells/ml, if concentration is > $5.0x10^4$ DEAD/ml dilute to below $5.0x10^4$ DEAD/ml

e. Distribute into 96-well plate, total of _____ wells (10 µl per well, DNA LoBind plates)

f. Add 3 µl 4x Dropseq Lysis Buffer + DTT

	4x Dropseq Lysis Buffer (final conc.)	Mix for 1 ml
50% Ficoll PM-400	24 % (6 % x4)	480 µl
20% Sarkosyl	0.8 % (0.2 % x4)	40 µl
0.5 M EDTA	80 mM (20 mM x4)	160 µl
1 M Tris pH 7.5	120 mM	120 µl
1 M DTT (add fresh)	200 mM (50 mM x4)	200 µl

g. Incubate 30 min at room temperature, shaking on vortexer (stage 4)

h. Fill up to 70 µl with water by adding 57 µl

i. Add 63 µl AMPure beads (0.9x) per well mix thoroughly (5x with pipette), incubate at room temperature for 5 min, bind to magnet, wash with 180 µl 80 % EtOH (take of 96-well plate from magnet, add EtOH, put on magnet, shift right, shift left, wait for beads to bind to magnet, take off EtOH)

j. Repeating washing with 180 µl 80 % EtOH, let beads air dry (at appearance of first crack, proceed with elution)

k. Elute in 40 µl water (add water, resuspend by pipetting, incubate 5 min at room temperature, put back on magnet, recover supernatant)

l. Repeat AMPure bead clean-up, add 32 µl beads (0.8x), elute in 15 µl water **POSSIBLE PAUSE POINT (FREEZE SAMPLES), but it is better to finish the SS synthesis**

Second Strand Synthesis – without TSO

2. Perform Second Strand Synthesis, 96 wells, add 5 µl of Mastermix to each well

	1x	X96+2	x____
10x Second Strand Buffer	2.0	196.0	
Second Strand Enzyme Mix (NEB E6111S)	0.8	78.4	
Ran. hexamers	0.1	9.8	
water	2.1	205.8	

Incubate 1:30 h at 16°C; Eppendorf Mastercycler 3D-seq -> Second Strand

3. Purify dsDNA by adding 20 µl AMPure XP beads (1.0x): add, mix, incubate 5 min at room temperature; put on magnet; wash twice with 180 µl 80 % EtOH; resuspend in 6 µl

4. Take out 1 µl from 5 random wells and use for qPCR to measure cDNA amount per well
   a. Ok to use qPCR stripes (applied Biosystems, MicroAmp strips + optical caps)

qPCR - 3D-seq measure cDNA amount per well

qPCR plate

	1	2	3	4	5
A	sample + primer 647/692		sample + primer 647/694		
B	water + primer 647/692		water + primer 647/694		
C					
D					
E					
F					
G					
H					

cDNA + 7.5 ul SYBR = 11.25 ul per well

expl.	wells	cDNA [ul]	water [ul]	SYBR [ul]
sample	4	4.40	12.10	33.00
water	4	4.40	12.10	33.00

Primer - 2.5 uM mix stock = 3.75 ul per well

647/692	binds in ligation barcode + 3'end of ActB
647/694	binds in ligation barcode + 3'end of Tubb3

to make 2.5 uM mix stock from 100 uM stock:

647	20 ul
692	20 ul
water	760 ul

- *qPCR measurement*: right machine -> New experiment from template -> Christin tagmentation standard.eds

- use “template_quantification_before_tagmentation.xlsx” for calculation of pg/µl

  _____ pg/µl x 6 µl (per well) x 96 wells = total mass _____ pg = _____ ng

  _____ pg/µl x 6 µl (per well) = _____ pg/well

  5 µl Amplicon Tagmentation Mix per 600 pg->_____ µl ATM per well (x pg/well * 5 µl/600 pg)

  PAUSE POINT (Freeze samples)

Tagmentation

5. Tagmentation: 5 min at 55°C, final volume 13 µl, add 7 µl to each well, mix well

	1x	X_____
Nextera TD Buffer (2x)	6.5	
Amplicon Tagmentation mix		
Sample	6.0	-----
Water		

6. Add 7 µl water to each well
7. Add 5 µl NT buffer to each well, mix well by pipette using 10 pipetting steps
8. Perform qPCR amplification for estimation (70 µl total volume, use 1 entire well for estimation with EvaGreen)

	1x
sample	25.0
10 µM N7xx (mine)	2.0
10 µM S5xx (kit)	2.0
5 µM oCS683	1.0
5 µM oCS684	1.0
20x Evagreen	3.5
water	2.6
Buffer	35.0
Terra polymerase	1.4

(600 pg should leave exponential phase at 12 cycles, 50 pg at 22 cycles; 552 pg was at 14-15 cycles in Exp_019)

Right machine -> New experiment from template -> Christin Terra PCR

- Determine optimal number of cycles for final PCR: change to linear view and take number of cycles where ΔRn is ~7500-10000 (wait for qPCR to finish!) **Figure 3**

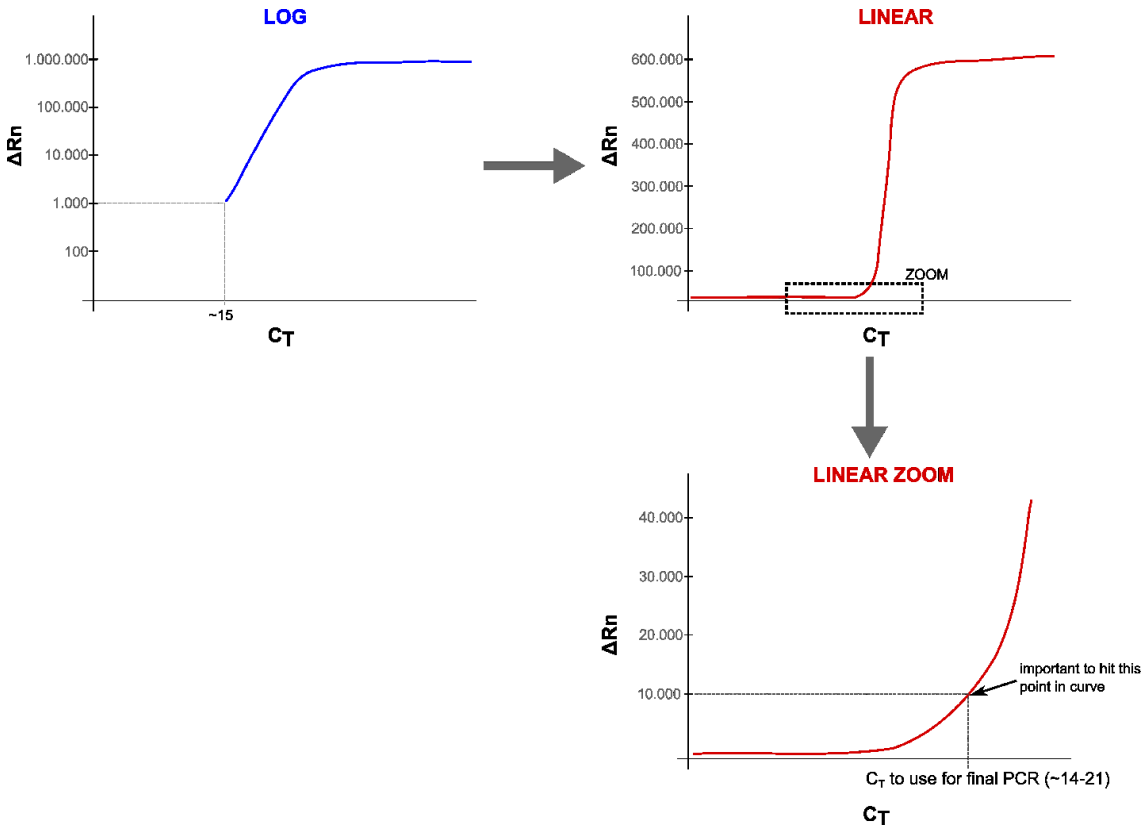


Figure 42: Determining the cycle number for the final PCR

9. Perform PCR, with Terra Polymerase for _____ cycles

	1x	X_____ – Terra
sample	25.0	-------
10 µM N7xx (mine)	2.0	--------
10 µM S5xx (kit)	2.0	-------
5 µM oCS683	1.0	
5 µM oCS684	1.0	
water	2.6	
Buffer	35.0	
Terra polymerase	1.4	

Primers used:

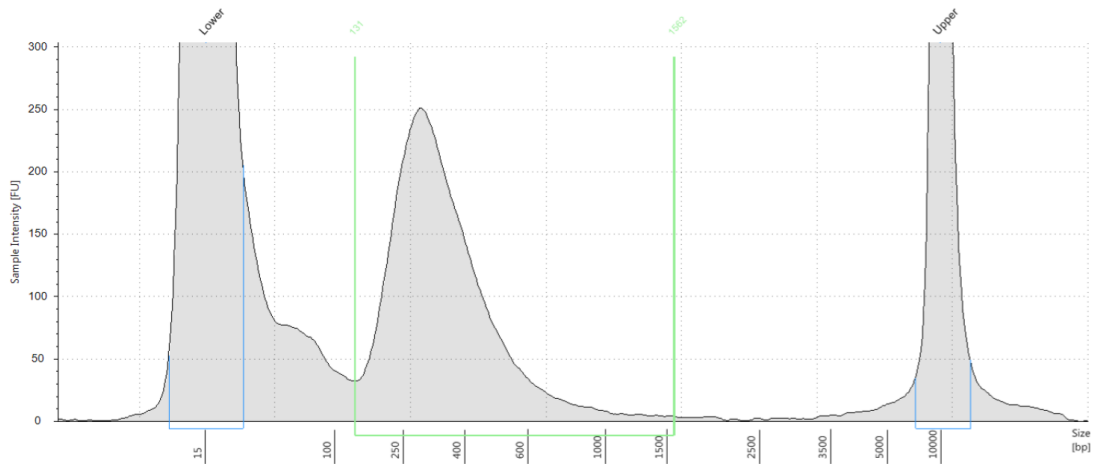
- N7____, N7____, N7____, N7____, N7____, N7____, N7____

Terra Polymerase:

98°C	2 min	
98°C	10 s	X cycles (check qPCR)
60°C	15 s	
68°C	60 s	
68°C	5 min	

10. Purify with AMPure beads: combine all wells w/ same PCR cycle in 2-4x 1.5 ml tubes (_____µl) and add ____ µl of AMPure XP beads (0.8x) to each tube, purify as usual and elute in 150 µl of water (use to combine all beads in one tube)

11. Repeat bead clean-up by adding _____ µl AMPure beads (0.8x), as usual and elute in 40 µl water
12. Tapestation profile and Qubit measurement
    a. Qubit should be between 0.5 – 2 ng/µl
    b. Tapestation profile example:



Region Table

From [bp]	To [bp]	Average Size [bp]	Conc. [pg/µl]	Region Molarity [pmol/l]	% of Total	Region Comment	Color
131	1562	361	250	1260	67.00		

13. For sequencing:
    a. use 1x150 nt Mid or HighOutput on NextSeq
    b. load 2-4 pM, 5 % PhiX: in the final dilution step according to Illumina's protocol use
        i. 130 µl 20 pM library (= 2 pM)
           1105 µl HT1
           65 µl 1.8 pM PhiX
        ii. 260 µl 20 pM library (= 4 pM)
           975 µl HT1
           65 µl 1.8 pM PhiX

    c. NO custom primers
    d. Read 1: 80 nt; Index 1: 8 nt; Index 2: 8 nt; Read 2: 67 nt

11 Selbstständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

_____

Christin Sünkel

Berlin, den

12 Publication list

Mol Cell. 2015 Jun 4;58(5):870-85. doi: 10.1016/j.molcel.2015.03.027. Epub 2015 Apr 23.

Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed.

Rybak-Wolf A*, Stottmeister C*, Glažar P*, Jens M, Pino N, Giusti S, Hanan M, Behm M, Bartok O, Ashwal-Fluss R, Herzog M, Schreyer L, Papavasileiou P, Ivanov A, Ohman M, Refojo D, Kadener S, Rajewsky N.

Citations: 353 (ISI Web of Knowledge – “Highly Cited Paper”) / 465 (Google Scholar) – 06.02.2019

Mol Cell. 2017 Apr 6;66(1):9-21.e7. doi: 10.1016/j.molcel.2017.02.021. Epub 2017 Mar 23.

Translation of CircRNAs.

Pamudurti NR*, Bartok O*, Jens M*, Ashwal-Fluss R*, Stottmeister C, Ruhe L, Hanan M, Wyler E, Perez-Hernandez D, Ramberger E, Shenzis S, Samson M, Dittmar G, Landthaler M, Chekulaeva M, Rajewsky N, Kadener S.

Citations: 166 (ISI Web of Knowledge – “Hot Paper” / “Highly Cited Paper”) / 237 (Google Scholar) – 06.02.2019

Cell Reports, in revision

A conserved circular RNA is required to keep neural cells in a progenitor state in the mammalian brain

Suenkel C, Cavalli D, Massalini S, Calegari F, Rajewsky N

In preparation

SPACE-seq: A method to preserve spatial information in single-cell RNA-seq

Suenkel C*, Karaiskos N*, Matz G, Rajewsky N

Patent application

Method for analyzing cell sample heterogeneity

Inventors: Suenkel C, Karaiskos N, Rajewsky N